Multivariate volatility modeling and forecasting with stable GARCH and Wishart autoregressive models

Dissertation
for the Faculty of Economics, Business Administration and Information Technology of the University of Zürich

to achieve the title of
Doctor of Economics

presented by

Matteo Bonato
from Italy

approved at the request of
Prof. Dr. Marc Paolella
Prof. Michael Wolf, Ph.D.
Prof. Loriano Mancini, Ph.D.

The Faculty of Economics, Business Administration and Information Technology of the University of Zürich hereby authorizes the printing of this doctoral thesis, without thereby giving any opinion on the views contained therein.
Zürich, October 21st, 2009

The Dean: Prof. Dr. Dr. J. Falkinger

To my Parents

Acknowledgements

In these four years passed in Zürich I had the fortune to share moments of happiness and to receive support in difficulties during the time I've spent writing this dissertation. I would like to thank my supervisor Marc Paolella and co-supervisor Michael Wolf for their guidance and useful comments that helped to improve the contents of this thesis. I am also deeply indebted with Marc for the knowledge, the new ideas and the teaching experience I acquired during the four years I served as his assistant. I would like to thank my officemate and friend Simon Broda, Yianna Tchopourian, Luca Taschini and Song Tao Wang for the fruitful discussions we had and above all, for being the great friends they are. A very special thank goes to my co-authors (in Manuscript 2) Massimiliano Caporin and Angelo Ranaldo for while working with them I really learnt to carry out high level research, to Loriano Mancini for sharing his advanced knowledge of high-frequency finance, and to Allan Timmermann for helping me to realize my dream of visiting the University of California in San Diego. To conclude, I would like to express my gratitude to my parents for believing in me and supporting my choice of doing a PhD; to my brothers Marco and Luca and my sister Marta for being always so close to me despite the geographical distance between us. And to Sandra for the love, patience and support she gave me.

Introduction

The enormous growth that financial markets have seen in the past 30 years has pushed academic research in the fields of economics, finance and statistics toward a better mathematical understanding and modeling of the dynamics by which markets are driven. One object in particular has captured the attention of scholars and financial practitioners: volatility.

The ability to model and predict volatility is in fact a key elements in three main fields of finance: risk management, asset allocation and option pricing.

The seminal paper that represents the cornerstone of the volatility modeling literature is without any doubt the one from the Nobel prize awarded Engle (1982). In his work, for the first time, a parametric model is proposed to model the volatility process in an autoregressive fashion. This model, called Autoregressive Conditional Heteroskedasticity (ARCH), assumes that the volatility (conditioned on the past information set) is latent and evolves as a function of the squared lagged residuals of the return process of the asset. This simple model is able to capture features typical of financial returns such as volatility clustering and excess of kurtosis. The academic literature that has been developed upon the ARCH model (and its generalization by Bollerslev, 1986 with the GARCH) consists probably of thousands of works. For an application to risk management one could, among all the others, see the work of Keuster et al. (2005). In an option pricing context a recent contribution was presented by Adesi et al. (2008).

A minor drawback of GARCH models is that volatility is assumed to be latent and a model for it needs to be estimated. They also use daily data so that the intra-day patterns are not captured. Although it represented a great improvement, this type of models do not generally provide satisfactory forecasts of volatility, with an R^2 of the regression of the forecasted variance on the squared daily returns bounded at 1/3 for a GARCH(1,1) model, see Andersen and Bollerslev (1998). In this paper it is also shown that GARCH models do provide accurate forecasts if the squared returns are replaced by the sum of intra-daily returns as measure of ex-post volatility. Thus, using information at a frequency higher than the daily it is possible to construct a more precise measure for the volatility.

Based on these last considerations and relying on the already existing theory of quadratic variation, Andersen et al. (2001a), showed that a consistent estimator of the daily integrated variance could be obtained by simply summing the squared intraday returns. This estimator of the volatility, now not latent but observable, is called *realized variance*. Unfortunately this estimator is biased in the presence of market micro-structure noise (infrequent trading, bid-ask bounce etc ...) and a lot of work has been done to present estimators for the integrated variance that are also robust to noise. Among the many proposed, see the two-time scale of Zhang et al. (2005), the range-based estimator of Martens and van Dijk (2007) and Christensen and Podolskij (2007) and the kernel estimators of Barndorff-Nielsen et al. (2008a). Application of realized volatility in risk management includes the works of Clements et al. (2008), Brownlees and Gallo (2008) and Giot and Laurent (2004). In option pricing see for example Bandi et al. (2007). In portfolio choice see Bandi et al. (2006) or de Pooter et al. (2006).

While most of the literature has focused in modeling univariate series of volatility, only recently efforts have been addressed toward the study of the dependence structure among assets. The ability of modeling jointly the dynamics of multiple assets and their correlation is crucial in portfolio optimization and risk evaluation. In my thesis I focused my attention on this issue, i.e. in modeling and forecasting (co)volatilities for multiple assets.

The reminder of this thesis is divided into two parts, consisting of one and two manuscript, respectively. Part I deals with multivariate generalized conditional autoregressive models. In Part II the object of investigation are the multivariate models for realized volatility with particular emphasis on a recently proposed model: the Wishart Autoregressive Model.

In particular, in Manuscript 1 a new multivariate volatility model is proposed. It combines the appealing properties of the stable Paretian distribution to model the heavy tails with the GARCH model to capture the volatility clustering. We assume that asset-returns follow a sub-Gaussian distribution, which is a particular multivariate stable distribution. In this way the characteristic function of the fitted returns has a tractable expression and the density function can be recovered by numerical methods. A multivariate GARCH structure is then adopted to model the covariance matrix of the Gaussian vectors underlying the sub-Gaussian system. The model is applied to a bivariate series of daily U.S. stock returns. Value-at-Risk for long and short positions is computed and compared with the one obtained using the multivariate normal and the multivariate Student's t distribution. Finally, exploiting the recent developments in the vast dimensional time-varying covariances modeling, possible feasible extensions of our model to higher dimensions are suggested and an illustrative example using the Dow Jones index components is presented.

In Manuscript 2, a joint work with Angelo Ranaldo (Swiss National Bank) and Massimiliano Caporin (University of Padua), we fucus on a new model for multivariate realized volatility. The increased availability of high-frequency data provides new tools for forecasting of variances and covariances between assets. However, recent realized (co)variance models may suffer from a 'curse of dimensionality' problem similar to that of multivariate GARCH specifications. As a result, they need strong parameter restrictions, in order to avoid non-interpretability of model coefficients, as in the matrix and log exponential representations. Among the proposed models, the Wishart autoregressive model introduced by Gourieroux et al. (2009) analyzes the realized covariance matrices without any restriction on the parameters while maintaining coefficient interpretability. Indeed, the model, under mild stationarity conditions, provides positive definite forecasts for the realized covariance matrices. Unfortunately, it is still not feasible for large asset cross-section dimensions. In this manuscript we propose a restricted parametrization of the Wishart Autoregressive model which is feasible even with a large cross-section of assets. In particular, we assume that the asset variances-covariances have no or limited spillover and that their dynamic is sector-specific. In addition, we propose a Wishart-based generalization of the heterogeneous autoregressive (HAR) model of Corsi (2009). We present an empirical application based on variance forecasting and risk evaluation of a portfolio of two US treasury bills and two exchange rates. We compare our restricted specifications with the traditional WAR parameterizations. Our results show that the restrictions may be supported by the data and that the risk evaluations of the models are extremely close. This confirms that our model can be safely used in a large cross-sectional dimension given that it provides results similar to fully parameterized specifications.

In Manuscript 3 an in-depth analysis of the estimation of the realized volatility Wishart Autoregressive model is presented. We focus in particular on the estimation of the degrees of freedom. A new estimator is proposed. Monte Carlo simulations show that this novel estimator is more efficient when compared to the standard estimator proposed in literature. We also show, again relying on simulation, that the presence of extreme observations in the variance-covariance process induces a bias toward zero of the estimated degrees of freedom, no matter which estimator one uses. However, the new proposed estimator is more robust compared to the standard one. To conclude, an empirical application to the S&P 500 - NASDAQ 100 futures realized variance-covariance series is carried out. It confirms that the estimated degrees of freedom, first, result sensitively lower when extremely high values in the volatility process are present and secondly, they increase with the sampling frequency.

Contents

Part I

GARCH models

Manuscript 1

Modeling fat tails in stock returns: a multivariate Stable-GARCH approach

1.1 Introduction

Volatility clustering, excess kurtosis and possible asymmetry are three universally recognized stylized fact typical of financial returns. Concerning the excess of kurtosis, building statistical models able to explain the presence of extreme observations is crucial, for example, in risk management. Extreme events are by definition rare and thus unexpected, and their impact in terms of financial loss is often underestimated.

In the finance literature, the presence of fat tails has often been ignored and many models rely on the multivariate Gaussian distribution as a building block. This choice is motivated by two reasons. On one side the central limit theorem provides a theoretical justification whenever the phenomenon of interest can be thought as the aggregation of a large number of micro-contributions. On the other side, the Gaussian distribution possesses a lot of useful properties that renders theoretical results easier to establish. The major shortcoming of using a Gaussian distribution is that it is incapable to model extreme events that in economic situations are reflected in extreme gains and losses and that are typical of financial markets. The multivariate Student's t and its skewed version (see Bauwens and Laurent, 2005 among others) represent valid alternatives but have the disadvantage of not being closed under summation. This implies a higher degree of difficulty when theoretical results need to be derived.

To capture the phenomenon of heavy tails, Mandelbrot (1963) and Fama (1965a) considered the family of stable Paretian distributions to model the unconditional distribution of financial returns. After their pioneering works, stable distributions have been investigated in many studies in the statistics and econometrics. Stable distributions in fact enjoy many of the properties of the Gaussian, such as closeness under summation, and a number of theoretical results in asset allocation and option pricing are available. See for example Fama (1965a), Fama (1965b), Ortobelli et al. (2002), Ortobelli and Rachev (2005), McCulloch (2003) and the survive by Bradley and Taqqu (2001).

In the academic literature, stable distributions have been proposed as a model for many types of physical and economic systems. There are several reasons for using a stable distribution to describe a system: the first is when, for theoretical reasons, we expect a non-Gaussian model, e.g. hitting times for a Brownian motion yielding a Lévy distribution; see Feller (1971) for this and other examples. The second reason is the Generalized Central Limit Theorem, which states that the only possible non-trivial limit of normalized sum of independent and identically distributed (i.i.d. hereafter) terms is stable. The third argument for modeling with stable distribution is empirical and related to the features of financial time series we presented before: heavy tails and skewness.

The classical objection against the stable assumption is that it has infinite variance. Empirical studies suggest the existence of the third or fourth moments for various financial data (cf. Pagan, 1996; Loretan and Phillips, 1994). To reach this conclusion, Hill (1975) or related tail estimators have been used, which are known to be not reliable for -even large- i.i.d. samples (see for example

Mittnik and Rachev, 1993; Mittnik et al., 1998; Paolella, 2001; Kratz and Resnick, 1996; Adler, 1997; McCulloch, 1997; Resnick, 1997) and even worse for data with GARCH structures (Kearns and Pagan, 1997). However, the reliability of the Hill estimator of the maximal moment exponent of a heavy tail distribution is not related to the i.i.d. nature of the marginal distribution, but to the question of whether the tails of the distribution are Paretian. In this case, in fact, the Hill estimator is effective whereas it can be biased if the tails are only asymptotically Paretian. Hence, the question of the maximally existing moments of financial return data is yet an open one. Similarly, arguing that a population is bounded and therefore must have a finite variance, should exclude the use of the normal distribution or other *ad hoc* distributions with infinite support as a model for the same population. As pointed out in Nolan (2003) the only justification provided is that the normal distribution gives a usable description of the shape of the distribution. The variance is one measure of spread, as the scale parameter in the stable case is another. Nowadays for many practitioners the variance is *the* measure of spread and any model with no finite variance is a priori rejected. If one would consider the variance just as the shape parameter of the Gaussian distribution, then the same can be done with the scale parameter of a stable distribution. When the matter of investigation is the shape of the distribution, the role the variance plays in the Gaussian case is played by the scale parameter in the stable case.

In this paper we give a time varying structure to the scale parameter of the distribution of the portfolio returns. This is done by modeling with a multivariate GARCH the covariance matrix of the Gaussian vectors we assume to be underlying the stable system.

Our goal is to model the joint - conditional and unconditional - distribution of the vector of asset returns of a portfolio assuming a multivariate stable distribution. If we consider the return at time t on a portfolio of, say, k assets, a univariate conditional (e.g. with a GARCH structure) or unconditional distribution can be fit to it. However, as the weight vector changes, the model has to be specified and fitted once again. If we work in a multivariate setting, the joint distribution of the returns can be directly used to compute the distribution of *any* portfolio.

One difficulty when working with stable distributions is that, in general, an analytical expression for the density function is not available, and such a class of distributions is defined only by its characteristic function (ch.f. hereafter). In the univariate setting one can use the inversion formula to recover the probability density function (p.d.f. hereafter). In this context the fast Fourier transform-based method (FFT) has been shown to perform particularly well when computing the density for a large number of data points (see Mittnik et al. (1999a)). Unfortunately in the multivariate case, the computation of the p.d.f. is even more complicated. A general expression for the characteristic function involves computing an integral with respect to the so called *spectral measure*, i.e. a finite measure Γ on the unit bell $\mathcal{S}_d \in \mathbb{R}^d$ with d being the dimension of the multivariate stable distributed vector.

Modeling the joint distribution of the asset-returns under the stable assumption is a challenging task due to the complexity of the expression for the ch.f. in the general case and the consequent estimation problems. However, under some assumptions, it is possible to transform the problem of the multivariate p.d.f. calculation to univariate p.d.f.'s calculations. We briefly present an overview of the present literature on the topic.

Nolan (2003) uses a multivariate stable elliptical distribution to model a multivariate series of financial returns. Given the particular expression of the ch.f. and exploiting the properties of the sub-Gaussian random variables, the parameters of the multivariate model are explicit functions of the parameters of the univariate series, which can be easily computed via ML estimation.

Lamantia et al. (2005) present an extension of the EWMA RiskMetrics model considering elliptically distributed returns and examine several new methods based on different stable distributional hypotheses of return portfolio. Finally, they discuss the applicability of temporal aggregation rules for each VaR and CVaR model proposed.

Doganoglu and Mittnik (2006) use a stable multi-index model to generate a multivariate stable system. The basic idea underling the multi-index model is that there exists a set of common market factors such that each return series evolves as a linear combination of the factors plus an additive idiosyncratic noise process that is independent of these factors. In this way the spectral measure is always discrete, and given the independence between the factors and the disturbance component, the (multivariate) p.d.f. can easily be calculated using univariate p.d.f.s. In Doganoglu et al. (2007) the same factor model is used and factors are considered to be conditionally varying. In both of the aforementioned papers, it is shown that the assumption of a multivariate (symmetric or asymmetric) stable distribution for the asset returns reduces the systematic bias in Value-at-Risk computation compared with the normal assumption.

In Garcia et al. (2006), the indirect inference method is adopted to estimate the parameters of an α-stable distribution. The skewed-t distribution is used as an auxiliary model. In Lombardi and Veredas (2009) the indirect inference method is extended to the multivariate case to estimate the parameters of elliptical stable distributions. This indirect estimation approach relies on the use of a multivariate Student's t distribution as auxiliary model. This distribution is also elliptical and the paper shows that its parameters have a one-to-one relationship with those of the elliptical stable. An application to 27 emerging markets stock indexes is also presented.

In this paper, we use sub-Gaussian random vectors to generate a multivariate stable system. Sub-Gaussian random vectors are a special case of stable random vectors. Some authors prefer the term "elliptically stable" as there are multiple meanings for sub-Gaussian in the probability literature, and they do not generally relate to stable distributions. We follow the notation in Samorodnitsky and Taqqu (1994) and their definition of sub-Gaussian random vectors. This choice allows us to have a tractable expression for the (multivariate) characteristic function and to express the scale parameter of the portfolio returns as a linear combination of the variances and covariances of the underlying Gaussian vectors. Under the sub-Gaussian hypothesis we are able to model the conditional and unconditional joint distribution of the asset returns. A multivariate GARCH model is introduced to describe the dynamics of the covariance matrix of the Gaussian vectors underlying the process. Given the computational complexity arising, we restrict our first analysis to two dimensions. The extension to a general d-dimensional case is theoretically straightforward, though computationally prohibitive. To circumvent this problem, we present some possible solutions that come from the vast dimensional covariances modeling literature. As done in Engle (2007) with the MacGyver estimator, a possible feasible way is to assume that the selected model (in this case the Dynamic Conditional Correlation of Engle, 2002) is correctly specified between every pair i and j and the parameters are obtained using simple aggregation procedures (such as median or mean) of the parameters estimated from all the bivariate pairs. A different approach is presented in Engle et al. (2008) where they construct a type of composite likelihood, which is then maximized to deliver the estimator. This composite likelihood is based on summing up the quasi-likelihood of subsets of assets.

The contribution of this paper is twofold. First, we use a multivariate stable distribution to model the joint distribution of asset returns. Stable distributions have tails fatter than the Gaussian distribution, thus in our formulation we take into account the phenomenon of excess kurtosis. Then, under the sub-Gaussian hypothesis, we impose a time varying structure for covariance matrix of the underlying Gaussian vectors. Given that the scale parameter of the distribution of the portfolio is a linear combination of the entries of this matrix, this originates a GARCH-type multivariate stable model: the multivariate stable GARCH model.

The second, indirect contribution of this paper is the extension, under the sub-Gaussian assumption, of the method in Mittnik et al. (1999a) to the bivariate case. This allows us to estimate the parameters of the model via maximum likelihood (ML). Our paper is the first, to our knowledge, that directly estimates all the parameters of a multivariate stable distribution in one step, i.e. without using univariate estimations as first step (this is the case of the projection method in Nolan, 2005, or the factor model of Doganoglu et al., 2006, 2007) or relying on indirect estimation (Lombardi and Veredas, 2009).

The rest of the paper is organized as follows. In Section 1.2, we present the family of stable distributions in the univariate and multivariate case. In Section 1.3, the sub-Gaussian hypothesis is used to model the joint distribution of the asset returns. In Section 1.4 the dataset used is described and results from the univariate estimations are reported. In Section 3.3, we explain the estimation procedure for the multivariate stable model. In Section 1.6, we provide an application of the model based on variance forecasting and risk evaluation of different portfolios. In Section 1.7 we introduce an extension of the stable-GARCH model to higher dimensions and show its feasibility using 29 stocks from the Dow Jones index. Section 1.8 concludes and proposes directions for future research. Proof of proposition 1 is reported in the Appendix.

1.2 Sub-Gaussian random vectors and their properties

The theory of univariate stable distributions was essentially developed in the 1920's and 1930's by Paul Lévy and Aleksandr Yakovlevich Khinchin. More recently, it was object of a monograph by Zolotarev (1986). This class of distribution nests two special distributions: the normal and the Cauchy. We now give two definitions of stable random variable.

Stable random variables do not possess a closed form for the p.d.f. and the distribution is defined via its ch.f. In literature, there are at least half a dozen different parameterizations. All involve different specifications of the ch.f. and are useful for various technical reasons. The parametrization most often used is the following.
The random variable X is said to have a stable distribution if there are parameters $0 < \alpha \leq 2, c > 0, -1 < \beta < 1$ and μ real such that its ch.f. has the form

$$\varphi_X(\theta) = \begin{cases} \exp\{-c^\alpha|\theta|^\alpha(1 - i\beta(\mathrm{sgn}\ \theta)\tan\frac{\pi\alpha}{2}) + i\mu\theta\} & \text{if } \alpha \neq 1, \\ \\ \exp\{-c|\theta|(1 + i\beta\frac{2}{\pi}(\mathrm{sgn}\ \theta)\ln|\theta|) + i\mu\theta\} & \text{if } \alpha = 1. \end{cases} \tag{1.1}$$

Since (1.1) is characterized by these four parameters, we will denote, as in Samorodnitsky and Taqqu (1994), stable distributions by $S_\alpha(c, \beta, \mu)$ and write $X \sim S_\alpha(c, \beta, \mu)$. We also write $X \sim S_\alpha S$ when X is symmetric α-stable, i.e. when $\beta = \mu = 0$. It is easy to see that in (1.1) when $\alpha = 2$ and $\alpha = 1$, the ch.f. coincides with that of the normal and Cauchy distribution, respectively.

The index α determines the thickness of the tails. When $\alpha = 2$, we have a Gaussian distribution. The smaller α, the fatter the tails become. For $0 < \alpha < 2$ the (fractional absolute) moments of $X \sim S_\alpha S$ of order α do not exist whereas for $\alpha = 2$ all positive moment exist. This indeed coincides with the special cases of the Cauchy ($\alpha = 1$) and the normal ($\alpha = 2$) distributions. Clearly the variance is not defined for any $\alpha < 2$.

In a multivariate setting the concept of stable random variable is replaced by stable random vectors. Stable random vectors possess (as the Gaussian random vectors) the appealing property that any linear combination of its components is indeed α-stable distributed. This can be a very useful property in portfolio theory as, under the assumption of a joint stable distribution for the asset returns, the returns of any portfolio of these assets is also α-stable distributed.

The expression for the c.f. given next involves an integration over $S_d = \{\mathbf{s} : ||\mathbf{s}|| = 1\}$, the unit sphere in $\mathbb{R}^d$. Observe that S_d is a $(d-1)$-dimensional surface. For example, S_1 is the two points set $\{-1, 1\}$ and S_2 is the unit circle. Let $0 < \alpha < 2$. Then $\mathbf{X} = (X_1, \ldots, X_d)$ is an α-stable random vector in $\mathbb{R}^d$ if and only if there exists a finite measure Γ on the unit sphere S_d of $\mathbb{R}^d$ and a vector μ^0 in $\mathbb{R}^d$ such that:
(a) If $\alpha \neq 1$,

$$\varphi_{\mathbf{X}}(\theta) = \exp\left\{-\int_{S_d} |(\theta, \mathbf{s})^\alpha|(1 - i\ sgn\ (\theta, \mathbf{s})\tan\frac{\pi\alpha}{2})\Gamma(d\mathbf{s}) + i(\theta, \mu^0)\right\}. \tag{1.2}$$

(b) If $\alpha = 1$,

$$\varphi_{\mathbf{X}}(\theta) = \exp\left\{\int_{S_d} |(\theta, \mathbf{s})|(1 + i\frac{2}{\pi}\ sgn\ (\theta, \mathbf{s}) \ln |(\theta, \mathbf{s})|)\Gamma(d\mathbf{s}) + i(\theta, \mu^0)\right\}. \tag{1.3}$$

The pair (Γ, μ^0) is unique. The vector $\mathbf{X}$ in is said to have spectral representation (Γ, μ^0). The measure Γ is called the *spectral measure* of the α-stable random vector $\mathbf{X}$.

To define the characteristic function of a symmetric α-stable random vector a necessary and sufficient condition is that $\mu^0 = 0$ and Γ is a symmetric measure on S_d (i.e. $\Gamma(Q) = \Gamma(-Q)$ for any Borel set Q in S_d).

A special case of symmetric α-stable random vectors is represented by the sub-Gaussian random vectors. For this class of vectors the spectral measure is always discrete and this results in a tractable expression for the characteristic function.

We start by presenting a useful characterization of the symmetric α-stable random variables. The following result shows that one can always transform a $S\alpha'S$ random variable into a $S\alpha S$ random variable, for any $0 < \alpha < \alpha'$.

Let $G \sim S_{\alpha'}(c, 0, 0)$ with $0 < \alpha' \le 2$ and let $0 < \alpha < \alpha'$. Let A be an α/α'-stable random variable totally skewed to the right with Laplace transform

$$E\left[\exp\{-\gamma A\}\right] = \exp\{-(2\gamma)^{\alpha/\alpha'}\}, \quad \gamma > 0,$$

i.e. $A \sim S_{\alpha/\alpha'}\left(2(\cos\frac{\pi\alpha}{2\alpha'})^{\alpha'/\alpha}, 1, 0\right)$, and assume G and A to be independent.
Then

$$X = A^{1/\alpha'} G \sim S_\alpha(c \cdot 2^{1/\alpha'}, 0, 0).$$

If we consider now the particular case where $\alpha' = 2$, then G becomes a zero mean Gaussian random variable; if the variance of G is σ^2 then we have, taking the different scaling convention for normal and stable random variables into account[1], $G \sim S_2(\sigma/\sqrt{2}, 0, 0)$. Thus,

$$X = A^{1/2} G \sim S_\alpha(\sigma, 0, 0).$$

This shows that every $S\alpha S$ random variable is conditionally Gaussian.

The previous results can be extended to random vector $\mathbf{X}$ as follows. Choose

$$A \sim S_{\alpha/2}\left(2(\cos\frac{\pi\alpha}{4})^{2/\alpha}, 1, 0\right) \tag{1.4}$$

[1]The random variable $X \sim N(0, \sigma^2)$ has ch.f. $\varphi_X(t) = \exp\{-\frac{1}{2}\sigma^2 t\}$. Using different notation we can write $X \sim S_2(c, 0, 0)$ and $\varphi_X(t) = -c^2t^2$ so that $c^2 = \sigma^2/2$ and we can express the scale parameter $c = \sigma/\sqrt{2}$

with $\alpha < 2$, so that its Laplace transform is

$$E[e^{-\gamma A}] = \exp\{-(2\gamma)^{\alpha/2}\}, \quad \gamma > 0. \tag{1.5}$$

Let $\mathbf{G} = (G_1, \dots, G_d)$ be a zero mean Gaussian vector in $\mathbb{R}^d$ independent of A. Then the random vector

$$\mathbf{X} = (A^{1/2}G_1, \dots, A^{1/2}G_d) \tag{1.6}$$

has a $S\alpha S$ distribution in $\mathbb{R}^d$ because, for any real numbers $b_1, \dots, b_d$ the linear combination $\sum_{k=1}^{d} A^{1/2}G_k = A^{1/2}\sum_{k=1}^{d} G_k$ is a $S\alpha S$ random variable and hence $\mathbf{X}$ is $S\alpha S$ and we write $\mathbf{X} \sim S_\alpha(\Sigma, 0, 0)$. (see Theorem 2.1.5 in for Samorodnitsky and Taqqu, 1994).

Definition 1. *(Samorodnitsky and Taqqu, 1994, p. 78) Any vector* $\mathbf{X}$ *distributed as in (3.D.2) is called a* sub-Gaussian *$S\alpha S$ random vector in $\mathbb{R}^d$ with underlying Gaussian vector* $\mathbf{G}$*. It is also said to be* subordinated *to* $\mathbf{G}$.

The sub-Gaussian symmetric α-stable random vector $\mathbf{X}$ defined in (3.D.2) has characteristic function

$$E[\exp\{i\sum_{k=1}^{d}\theta_k X_k\}] = \exp\{-|\sum_{i=1}^{d}\sum_{j=1}^{d}\theta_i\theta_j\sigma_{ij}|^{\alpha/2}\}, \tag{1.7}$$

where $\sigma_{ij} = E[G_iG_j]$ and $\sigma_i^2 = E[G_i^2]$, $i, j = 1, \dots, d$, are the covariances and variances of the underlying Gaussian random vectors $(G_1, \dots, G_d)$.

When working in a multivariate setting, the ability to define (and thus to model) the dependency structure among assets is of fundamental importance. The covariance function is an extremely powerful tool in the study of Gaussian random elements, but it is not defined when $\alpha < 2$. The *covariation* is designed to replace the covariance when $1 < \alpha < 2$. Unfortunately, it lacks in some of the desirable properties of the covariance. For (X_1, X_2) jointly $S\alpha S$, $\alpha > 1$ and the spectral measure Γ, the covariation of X_1 on X_2 is the real number

$$[X_1, X_2]_\alpha = \int_{S_2} s_1|s_2|^{\alpha-1}\mathrm{sgn}(s_2)\Gamma(d\mathbf{s}) \tag{1.8}$$

When X_i and X_j are sub-Gaussian random vectors, then entries of the covariation matrix posses the closed-form expression:

$$[X_i, X_j]_\alpha = \sigma_{ij}\sigma_j^{(\alpha-2).} \tag{1.9}$$

Notice that $[X_i, X_j]_\alpha = [X_j, X_i]_\alpha$ if $\sigma_i^2 = \sigma_j^2$, i.e. the covariation between two α-stable random variables is generally not symmetric in its arguments.

1.3 The model

Of primary importance when working with multiple assets, it is to consider the dependence structure among them. After Engle's (1982) seminal paper, ARCH and GARCH models have been extended to the multivariate case in many different ways, so that the covariance structure among assets be time-varying. Stable random vectors, however, do not possess a covariance matrix and it would seem reasonable to replace the covariance matrix with the covariation matrix and model it with a GARCH structure. However, the entries of the covariation matrix are non-linear functions of the covariances of the underlying Gaussian vectors generating the sub-Gaussian vectors and thus do not have a direct interpretation. An other disadvantage is the non-symmetry of the matrix, that leads to an even less clear understanding of its meaning. We show now that under the sub-Gaussian hypothesis, the scale parameter of the distribution of the portfolio returns is a linear function of the covariance matrix of the underlying Gaussian vectors. Thus, heteroskedasticity can be introduced by assuming a GARCH specification of this covariance matrix.

Let A be a totally skewed α-stable random variable as defined in (3.D.1), $0 < \alpha < 2$. Let $G_t = (G_{1t}, \ldots, G_{Nt})$ be a conditionally zero mean Gaussian vector independent of A for every $t = 1, \ldots, T$, i.e.

$$G_t|\mathbb{I}_{t-1} \sim N_N(\mathbf{0}, \Sigma_t), \quad \Sigma_{ij,t} = \sigma_{ij,t}, \quad i,j = 1, \ldots, N, \quad t = 1, \ldots, T,$$

where $\sigma_{ij,t} = E[G_{it}G_{jt}]$ and $\mathbb{I}_{t-1}$ is the information available at time $t-1$. Consider the vector of asset returns at time t, r_t. Define $\epsilon_t = r_t - \mu_t$ the vector of demeaned returns.

Assumption 1. *The vector*

$$\epsilon_t = (\epsilon_{1t}, \ldots, \epsilon_{Nt}), \quad t = 1, \ldots, T,$$

is a sub-Gaussian $S\alpha S$ random vector with underlying Gaussian vector $\mathbf{G}_t$; i.e.

$$\epsilon_t = (A^{1/2}G_{it}, \ldots, A^{1/2}G_{Nt}), \quad t = 1, \ldots, T.$$

Under Assumption 1 we are able to define the distribution of the portfolio returns.

Proposition 1. *Let $r_t = (r_{1t}, \ldots, r_{Nt})$ be a $N \times 1$ sub-Gaussian vector of asset returns. Denote by P_t the return of the portfolio at time t, i.e. $P_t = \sum_{i=1}^N \omega_i r_{it}$ with $\omega = (\omega_1, \ldots, \omega_N)$ representing the portfolio weights. Then:*

$$P_t|\mathbb{I}_{t-1} \sim S_\alpha(\sigma_t, 0, \omega'\mu_t), \tag{1.10}$$

with $\sigma_t^2 = \omega'\Sigma_t\omega$.

See the Appendix at the end of the paper for the proof. Note that α is not the characteristic exponent of the distribution of the portfolio returns, but the characteristic exponent of the multivariate stable distribution of the asset returns. Estimating α directly from the distribution of P_t

would be the erroneous procedure. The stability index, in fact, determines the joint distributions of the asset-returns and does not depend on the way the portfolio is constructed.

Proposition 1 tells that heteroskedasticity in the model (in terms of time varying scale parameter σ_t) can be introduced by simply allowing the covariance matrix of the Gaussian random vector, Σ, to be time varying. A GARCH structure to model the dynamics of Σ_t originates what we define a *multivariate Stable GARCH model*.

Next section introduces the dateset used and preliminary results of the univariate estimation are presented. Then, the multivariate estimation procedure is described, starting from most challenging issue: the calculation of the $S\alpha S$ pdf.

1.4 Data description and univariate estimations

We model daily return data from the Procter & Gamble (PG) and Merck & Co. (MRK) stock using a sample from January 2, 1990 to May 7, 2007 implying 4373 observations obtained from Yahoo! Finance. Continuously compounded percentage returns are considered, i.e. daily returns are measured by log-differences of closing pricing multiplied by 100.

Sample path for prices and returns and marginal kernel density estimates are given in Figure 1.1. From the plot of the returns we notice the presence of extreme events for both series which reflects in consistent losses in terms of returns.

For the first series, Procter & Gamble, the crash-down happened the 7th of March 2000. It followed the announcement that its earnings for the next two quarters would have been lower than expected. The title tumbled 31 percent that day, one of the biggest single day losses ever for a company listed on the Dow Jones Industrial Average.

The crash of Merck & Co. is related to its arthritis drug Vioxx. On September 30, 2004, the company announced that it was immediately withdrawing Vioxx from world markets after a data safety monitoring board, overseeing a long-term study of the drug, recommended that the study be halted due to an increase risk of serious cardiovascular events among members of the study group. The company's abrupt decision to withdraw Vioxx contradicted its prior public announcement repeatedly touting the safety of Vioxx. Following the withdrawal of Vioxx from the markets, Merck's stock price immediately plummeted by 26 percent, resulting in billions of dollars in losses for the investors.

Table 1.1: Descriptive statistics of the Procter & Gamble - Merck.

	Correlation PG	matrix MRK	Mean	Skewness	Kurtosis
PG	1	0.31	0.05	-2.80	67.20
MRK	0.31	1	0.04	-1.31	27.32

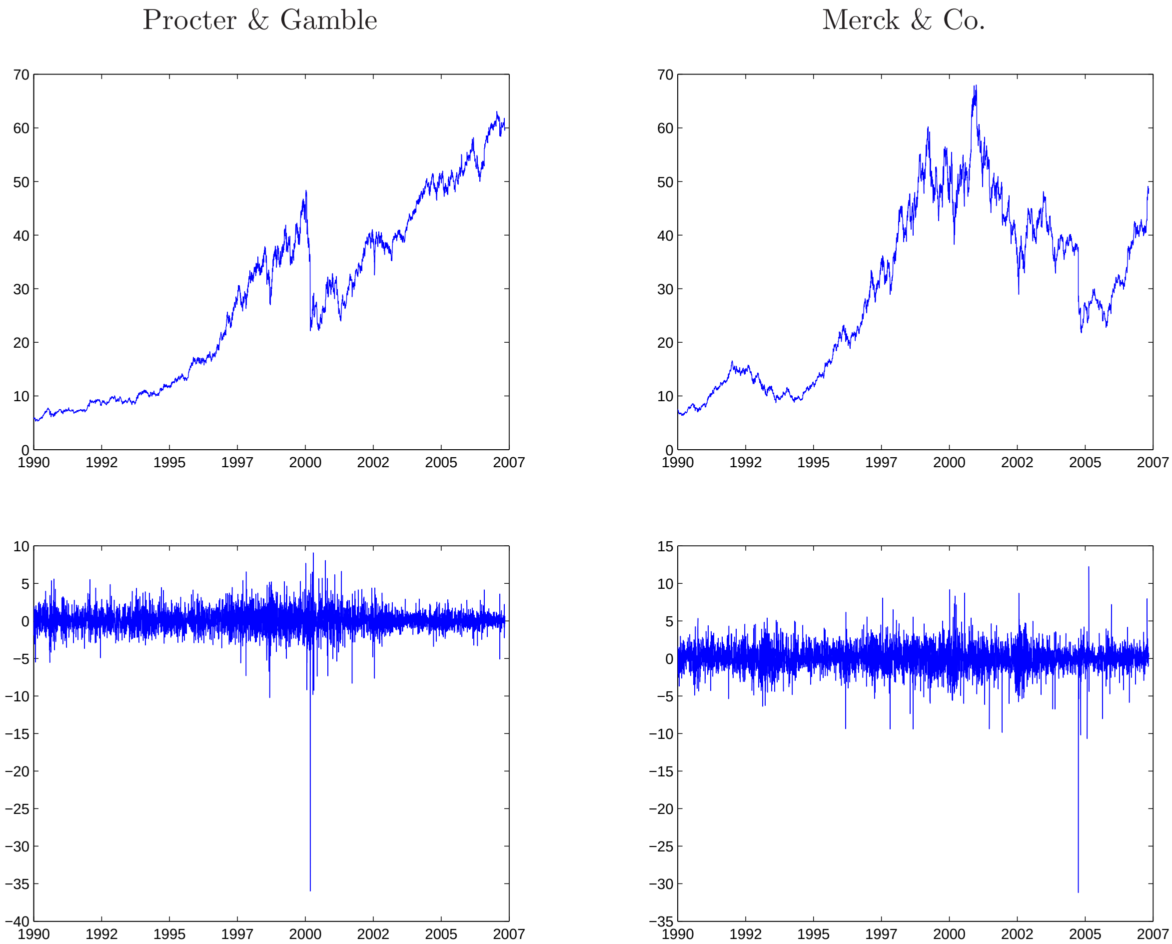


Figure 1.1: Procter & Gamble (left) and Merck & Co. (right) in levels (top) and daily returns (bottom)

Descriptive statistics for the univariate series are given in Table 1.1. From a preliminary exploratory analysis we see that both series are extremely leptokurtic and present asymmetry. Sample kurtosis are 67.2 and 27.3. This clearly indicates we are far away from a normal distribution for the returns. Both series present an index of skewness which differs from zero, -2.8 and -1.31, respectively. Table 1.2 reports the estimates of the parameters when a univariate stable model is fitted to the series. Parameters are estimated via maximum likelihood. See Mittnik et al. (1999b) for a detailed exposition of the estimation procedure. In the family of stable distributions the thickness of the tails is captured by the the estimated value of α, which, for both series, is statistically different from 2, the stability index of a Gaussian distribution. This confirms that the normal assumption for the returns does not seem appropriate.

An interesting issue in this basic analysis involves the presence or not of asymmetry in the series. In the class of stable distribution the asymmetry is regulated by the parameter β. In our two series the estimates of β are not statistically different from zero at level 1% and the hypothesis of asymmetry is rejected. To confirm these, we also computed the likelihood ration test (LRT)

Table 1.2: Results for the unconditional univariate asymmetric α-stable model. Standard errors are expressed in parenthesis and were computed using the numeric approximation of the hessian matrix.

	$\hat{\alpha}(\widehat{SE})$	$\hat{c}(\widehat{SE})$	$\hat{\beta}(\widehat{SE})$	$\hat{\mu}(\widehat{SE})$
PG	1.74	0.87	0.13*	0.07
	(0.0243)	(0.0132)	(0.0601)	(0.0222)
MRK	1.78	1.03	0.0016	0.05
	(0.0234)	(0.0153)	(0.0093)	(0.0232)

* significant at level 95%. No star means the parameter is not statistically significant at level 95% (for β and μ only).

for the asymmetric against the symmetric model. The values of the test statistic is 3.59 and 0 for the Procter and Gamble and Merck and Co., respectively. As for this test, the critical value at level 5% is 3.84, we cannot reject the hypothesis of symmetry for the two series at level 5% and lower. This is not in line with the negative values of the skewness indexes. To explain these apparently contradictory results, two points must be made. First, under the assumption that a sample comes from a Stable distribution, skewness and kurtosis theoretically do not exist, and thus should not be used by any mean as indicators of asymmetry and fatness of the tails. Second, as widely documented literature (see Kim and White, 2004, for a review alternative measures), skewness and kurtosis are not robust to the presence of extreme observations.

We suspect that the single extreme event that characterizes the two series in this study leads to erratic conclusions regarding the presence of asymmetry. Checking our hypothesis is straightforward. We simply re-estimated the parameters without considering, only in this robustness analysis, the observations corresponding to the crashes of the 7th of March 2000 for Procter & Gamble and of the 30th of September 2004 for Merck & Co . The result is surprising. Table 1.3

Table 1.3: Descriptive statistics of the Procter & Gamble - Merck after the outliers have been removed.

	Correlation	matrix	Mean	Skewness	Kurtosis
	PG	MRK			
PG	1	0.24	0.06	-0.14	7.17
MRK	0.24	1	0.05	-0.11	6.70

reports the descriptive statistics. Removing only one extreme observation per sample lead the skewness value from -2.80 to -0.14 for Procter & Gamble and from -1.31 to -0.11 for Merck& Co. The index of kurtosis was also notably affected and its new value is considerably lower now that these two big losses have been removed. Skewness and kurtosis are usually coarse preliminary measure of the asymmetry and of the fatness of the tails of a distribution when the model in mind for a given data set is the normal distribution. One question that immediately arises after the outliers removal is: if skewness and kurtosis dramatically changed after only one observation has been removed, what happens to the parameters of the stable distributions? In particular,

Table 1.4: Results for the unconditional univariate asymmetric α-stable model after the outliers have been removed. Standard errors are expressed in parenthesis and were computed using the numeric approximation of the hessian matrix.

	$\hat{\alpha}(\widehat{SE})$	$\hat{c}(\widehat{SE})$	$\hat{\beta}(\widehat{SE})$	$\hat{\mu}(\widehat{SE})$
PG	1.74	0.87	0.13	0.07
	(0.0243)	(0.0133)	(0.0934)	(0.0327)
MRK	1.78	1.03	0.008	0.06
	(0.0234)	(0.0153)	(0.0193)	(0.0277)

* significant at level 95%. No star means the parameter is not statistically significant at level 95% (for β and μ only).

did this affect the tail index α and the asymmetry parameter β? We answer to this question analyzing the estimates for the unconditional distribution of the two series of asset returns. The estimates are reported in Table 1.4. The values of α and β are for the two series are not different and confirm the previous conclusions: no clear asymmetry and no thin tails.

We move now to the model the conditional distribution of the sample. We stress that the extreme observation present in each series has not been discharged as, first, they do justify the use of a Stable distribution, second, dropping data means losing information that will affect, for example, any Value-at-Risk measurement exercise.

The presence of serial autocorrelation between the returns as well as the phenomenon of heteroskedasticity are well documented facts in literature. This last aspect gains more importance under the stable hypothesis, in which one can claim that the fatness of the tails is a phenomenon associated to the presence of heteroskedasticity and not concerning the distributional assumption for the standardized residuals.

To remove the presence of serial autocorrelation in the series we fit an AR(1) structure in the r_t:

$$\Phi(L)(r_t - \mu) = \epsilon_t \tag{1.11}$$

where $\Phi(L) = 1 - \phi L$ in an AR lag polynomial of order 1. The conditional mean of r_t, μ_t is defined to be $\mu + \phi(r_t - \mu)$.

A GARCH-type structure is then introduced to model the conditional variance of the residuals. It reads

$$\epsilon_t = \sigma_t \, z_t, \tag{1.12}$$

$$\sigma_t = f(\sigma_{t-1}, \ldots, \sigma_1, \epsilon_{t-1}, \ldots, \epsilon_1). \tag{1.13}$$

If the model is correctly specified, under the stable assumption, the standardized residuals are independent and identically distributed following a stable distribution with common tail index α, zero location parameter and time varying scale parameter σ_t. One property of the family of stable Paretian distribution is the so called *stability under summation* , or, in short, SuS. This means that the sum of observations of consecutive non-overlapping subsamples of length S the estimate of the stability index does not move toward 2 as S increases. A common approach to testing wether several financial return series exhibit summability is to estimate their respective tail index, α , at the daily, weekly and monthly aggregate levels and then, for each level of aggregation, informally compare $\sum \mathbb{I}_{(0,2)}[\hat{\alpha} + 2\hat{SE}(\alpha)]$. i.e. the number of $\hat{\alpha}$ plus two standard errors which are below the threshold 2.0 (see Akgiray and Booth, 1988 and Akgiray et al., 1989). The obvious criticism of not using i.i.d. data notwithstanding, these studies do not take the small sample properties of the vector $\hat{\alpha}(s)$ into consideration, where $\hat{\alpha}(s)$ denotes the estimates of α for given level of aggregation s. For a given i.i.d. stable Paretian series of length T, Paolella (2001) proposes examining the extent to which $\hat{\alpha}(s)$ changes with respect to s. For a non-stable i.i.d series, $\hat{\alpha}(s)$ is expected to increase with s, which, to the first order, can be approximated by a linear trend. The proposed summability test estimates this relation as a linear regression of $\hat{\alpha}(s)$ onto a constant and vector s and considers the behavior of the latter's coefficient, say $\hat{b}$. The test takes the form

$$\tau_T(\alpha) = \frac{\hat{b}}{\widehat{SE}(\hat{b})}$$

and the cutoff values under the null of stable Paretian data have been computed, for given T and level γ, as smooths functions of α.

As pointed out in Mittnik et al. (2000), the rejection of a SuS-based test for the stable Paretian hypothesis for asset returns could be the consequence of the fact that the series is not stable distributed or of the presence of serial dependence, or, in fact, both. When an ARMA-GARCH filter is applied to financial returns, the residuals will be much closer to i.i.d. than the unfiltered counterpart. Nevertheless, few would insist that observed data series are really generated by an ARMA-GARCH process. Thus, SuS based tests may be jeopardized by the extent to which the filtered residuals deviate from i.i.d.ness. Although attaining genuine i.i.d. residuals is an ideal which will hardly be reached, it is essential that one uses models that filter the date as effectively as possible.

To this end, among the numerous extension of the standard GARCH specification, we decide to adopt the Quadratic GARCH, or Q-GARCH model, introduced by Sentana (1991, 1995). This model allows current conditional volatility to react asymmetrically to the size of the previous periods' innovations, which is necessary to capture the well-know leverage effect as originally noted by Black (1976).

A generalization of the Q-GARCH is represented by the Stable Q-GARCH(p, q) model, introduced in Mittnik et al. (2000), which reads:

$$\sigma_t^\delta = c_0 + \sum_{i=1}^{p} c_i \epsilon_{t-i} + \sum_{i=1}^{p} c_{ii} |\epsilon_{t-i}|^\delta + 2 \sum_{i=1}^{p} \sum_{j=i+1}^{p} c_{ij} \epsilon_{t-i} \epsilon_{t-j} + \sum_{j=1}^{q} d_j \sigma_{t-j}^\delta. \tag{1.14}$$

This model, denoted $S_{\alpha,\beta}^\delta$ Q-GARCH (p, q), will be useful in case where the second moment of ϵ_t

Table 1.5: Results for the conditional univariate asymmetric α-stable model. Standard errors are expressed in parenthesis and were computed using the numeric approximation of the hessian matrix

Returns.	$S_{\alpha,\beta}$ Q-GARCH			
series	$\hat{\alpha}(\widehat{SE})$	$\hat{\beta}(\widehat{SE})$	$\hat{V}_Q(\widehat{SE})$	$\tau_T(\alpha)$
PG	1.864	0.0757	0.993	1.13
	(0.0183)	(0.0990)	(0.0049)	
MRK	1.851	-0.0160	0.985	1.64
	(0.0196)	(0.0942)	(0.0028)	

[a]Column $\tau_T(\alpha)$ is the summability test statistic; no sign means we cannot reject the null of stability at the 90 % (and then also the 95 % and the 99 %).
* significant at level 95%. No star means the parameter is not statistically significant at level 95% (for β only).

does not exist, e.g. with non-normal stable Paretian innovations.

The solutions of the process are strictly stationary when

$$V_Q = \sum_{i=1}^{p} c_{ii} E|z_t|^{\delta} + \sum_{j=1}^{q} d_j. \tag{1.15}$$

A closed form expression for $E|z_t|^{\delta}$ is presented in Mittnik et al. (2002). When $\delta = 2$ and the absolute values are replaced by the squared value the model becomes the original Q-GARCH model of Sentana (1991, 1995). As noted in Haas et al. (2005) for a variety of data sets and in agreement with the findings in Panorska et al. (1995), restraining δ to be one results in a very little loss of goodness of fit. Thus we restrict ourselves to the case $\delta = 1$.

Selected parameter estimates for the two series are reported in Table 1.5. The stable tail index $\hat{\alpha}$ is 1.86 for Procter & Gamble and 1.85 for Merck & Co. As expected, these values are higher than the ones obtained when fitting the unconditional distribution, as some of the fatness of the tails has been captured by the Q-GARCH filter. In both series the estimates of the the skewness parameter $\hat{\beta}$ are not statistically significant, implying symmetry in the distributions. The fourth column reports the estimates of the persistence measure $\hat{V}_Q$. Column $\tau(\alpha)$ is the summability test statistics for the residuals as proposed in Paolella (2001). The stability test delivers both the statistics $\tau_T(\alpha)$ and the appropriate cutoff values (as function of the estimated stable tail index α and sample size T) at the 90, 95 and 99 percent level. The Procter & Gamble Q-GARCH residuals yield $\tau_T(\alpha) = 1.13$. For Merck & Co. its value is $\tau_T(\alpha) = 1.64$. The nully hypothesis of stability cannot be rejected at the 90 (or higher) percent level.

1.5 Multivariate model estimation

This preliminary analysis suggests that a stable Paretian model for the joint distribution of the returns seems appropriate. Our next step is hence to fit a multivariate sub-Gaussian model to the unconditional (and conditional) joint distribution. We start presenting the multivariate extension of the method of Mittnik et al. (1999a) we use to recover the p.d.f.s necessary to construct and maximize the likelihood of the returns under the sub-Gaussian assumption.

1.5.1 Efficient calculation of the $S\alpha S$ PDF's

As has been said previously, there is no closed form expression for the density of a multivariate stable distribution. Only the characteristic function is known. When working with sub-Gaussian random vectors the ch.f. has a tractable expression. Starting from the ch.f. it is possible, via the inversion formula, to recover the multivariate p.d.f.. To fulfill this we extended to the bivariate setting the method of Mittnik et al. (1999a).

In the univariate case, the $S\alpha S$ p.d.f. can be written as

$$f(x;\alpha,\sigma,\mu)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-ixt}\varphi(t)dt. \tag{1.16}$$

which is the inversion formula to recover the p.d.f. when the ch.f. is known.

In the bivariate case the formula becomes

$$f(x,y;\alpha)=\left(\frac{1}{2\pi}\right)^2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-ixt_1-iyt_2}\varphi(t_1,t_2)dt_1dt_2. \tag{1.17}$$

In the univariate setting Mittnik et al. (1999a) derives the p.d.f. directly as the Fourier transform of the ch.f. in (1.1). We extend their method to the multivariate case. For ease of notation, and because it is what we in practice implemented, we restrict our analysis to the bivariate case.

The FFT is an efficient way of computing the Fourier transform. The integral in (1.17) will be calculated in a lattice of $N_1\times N_2$ equally-spaced points with distance h_1 and h_2, namely $x_k=(k-1-(N_1/2))h, k=1,\dots,N_1$ and $y_l=(l-1-(N_2/2))h, l=1,\dots,N_2$. Letting $t_1=2\pi u, t_2=2\pi v$,(1.17), becomes

$$f\left(\left(k-1-\frac{N_1}{2},l-1-\frac{N_2}{2}\right)\right)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varphi(2\pi u,2\pi v)e^{-i2\pi u(k-1-\frac{N_1}{2})h_1-i2\pi v(l-1-\frac{N_2}{2})h_2}dudv. \tag{1.18}$$

The double integral in (1.18) can be approximated by using the rectangle rule for the lattice created with the $N_1\times N_2$ points with spacing s_1 and s_2, i.e.

$$f\left(\left(k-1-\frac{N_1}{2},l-1-\frac{N_2}{2}\right)\right)\approx s_1s_2\sum_{n_1=1}^{N_1}\sum_{n_2=2}^{N_2}\varphi\left(2\pi s_1(n_1-1-\frac{N_1}{2}),2\pi s_2(n_2-1-\frac{N_2}{2})\right)\times \tag{1.19}$$
$$\exp\left\{-i2\pi(n_1-1-(N_1/2))(k-1-(N_1/2))h_1s_1-i2\pi(n_2-1-(N_2/2))(l-1-(N_2/2))h_2s_2\right\}.$$

By setting in (1.19) $s_1=(h_1N_1)^{-1}$ and $s_2=(h_2N_2)^{-1}$, we obtain the approximation

$$f\left(\left(k-1-\frac{N_1}{2},l-1-\frac{N_2}{2}\right)\right)\approx s_1(-1)^{k-1-(N_1/2)}s_2(-1)^{l-1-(N_2/2)}\times \tag{1.20}$$

$$\sum_{n_1=1}^{N_1}\sum_{n_2=2}^{N_2}(-1)^{n_1+n_2-2}\varphi\left(2\pi s_1(n_1-1-\frac{N_1}{2}),2\pi s_2(n_2-1-\frac{N_2}{2})\right)\times$$
$$\exp\left\{(-i2\pi(n_1-1)(k-1))/N_1-i2\pi(n_2-1)(l-1))/N_2\right\}.$$

The summation in (1.20) is computed by applying FFT to the sequence

$$(-1)^{n_1+n_2-2}\varphi\left(2\pi s_1(n_1-1-\frac{N_1}{2}),2\pi s_2(n_2-1-\frac{N_2}{2})\right).$$

The (k^{th}, l^{th}) element of the resulting sequence is normalized by

$$s_1(-1)^{k-1-(N_1/2)}s_2(-1)^{l-1-(N_2/2)}$$

to obtain the pdf value for each lattice point.

Along with the method proposed by Mittnik et al. (1999a), the procedure to obtain p.d.f. values for irregularly-spaced data consists of two steps. First, we specify a lattice using two equally spaced grids covering the range of data and compute the p.d.f. on the lattice of points. This is done using the Matlab function `fft2` to compute the two dimensional Fourier transform on the $N_1 \times N_2$ matrix of the ch.f. calculated on the lattice. In the second step we use two dimensional linear interpolation to the data points falling between the lattice values. To accomplish this we used the Matlab function `interp2.` Mittnik et al. (1999a) suggest that for $1.6 < \alpha < 1.9$, which are values typically of financial data, setting $h = 0.01$ and $N = 2^{13}$ leads to a fast and sufficient accurate approximation. Unfortunately, using the same values for these tuning parameters in the bivariate case, increases in a notable way the computational burden. In this case greater speed is more desirable than greater accuracy. Therefore we set $h_1 = h_2 = 0.04$ and $N_1, N_2 = 12$.

1.5.2 Maximum likelihood estimation

Under Assumption 1, we defined the error term as

$$\epsilon_t = r_t - \mu_t = (A^{1/2}G_{1,t}, \ldots, A^{1/2}G_{N,t})$$

where A is a totally skewed α-stable random variable and G_t is a conditionally zero mean Gaussian vector independent of A with $G_t \sim N_N(0, \Sigma)$. Let $\theta = (\alpha, vech(\Sigma))'$ the vector of unknown parameters to be estimated and $vech(\Sigma)$ is the operator that stacks the upper triangular components of the matrix Σ. The ML estimate of θ is obtained by maximizing the log-likelihood function

$$
\begin{aligned}
l(\theta,\epsilon) &= \sum_{t=1}^{T} \log f(\epsilon_t;\theta) \\
&= -\frac{T}{2}\log(|\Sigma|) + \prod_{i=1}^{T} f(z_t;\alpha)
\end{aligned}
\tag{1.21}
$$

where $z_t = \Sigma^{-1/2}\epsilon_t$ is distributed as $S_\alpha(\mathbf{I}_N, 0, 0)$, $\mathbf{I}_N$ denotes the $N \times N$ identity matrix and the joint density function f is obtained using the method proposed in Section 1.5.1.

In the univariate case, Mouchel (1973) investigates the theoretical properties of the ML estimator for θ and shows its asymptotical normality under certain regularity conditions, i.e.

$$
\sqrt{T}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N(0, I^{-1}(\theta_0))' \tag{1.22}
$$

where "$\xrightarrow{d}$" stands for convergence in distribution and I denotes the Fisher information matrix

$$
I(\theta_0) = -E\left(\frac{\partial l(\theta;\epsilon)}{\partial\theta\partial\theta'}\right), \tag{1.23}
$$

which can be approximated either using the Hessian matrix arising in the maximization, or as in Nolan (1997), by numerical integration. In our ML estimation algorithm, we maximize log-likelihood function (1.21) numerically. Rather than employing constrained optimization, we follow Mittnik et al. (1999b) and estimate a transformed version of θ, say $\tilde{\theta} = (\tilde{\alpha}, \tilde{\sigma}_{11}, \tilde{\sigma}_{12}, \tilde{\sigma}_{22})'$, such that $\theta = h(\tilde{\theta})$. The transformation we adopt takes the form:

$$
\alpha = 1 + \frac{1}{1+\tilde{\alpha}^2}; \quad \sigma_{11} = \tilde{\sigma}_{11}^2, \quad \sigma_{22} = \tilde{\sigma}_2^2, \quad \sigma_{12} = \tilde{\sigma}_{12}. \tag{1.24}
$$

The transformation of α ensures that it assumes values in (1,2] as in many application it is assumed that the first moment of ϵ_t exists. The transformation of the diagonal elements of Σ rules out negative values for the variance. With the parameter transformations in place and defining the gradients $\delta_{\tilde{\theta}} h = \frac{\partial h}{\partial \tilde{\theta}'}$, (1.22) becomes

$$
\sqrt{\hat{\theta} - \theta_0} \xrightarrow{d} N\left(0, \Delta_{\tilde{\theta}} h I^{-1}(\tilde{\theta_0}) \Delta_{\tilde{\theta}} h'\right). \tag{1.25}
$$

To investigate the performance of this FFT-based ML we conduct a Monte Carlo study. In view of the complexity of the algorithm and of the computational burden required, we fucus only on values for the stability index that are compatible with the ones generally found in the financial literature. Recall also that the estimated stability index usually assumes higher values when the model is fitted to the series of residuals of GARCH filter applied to eliminate the heteroskedastic component. So, on the basis of these previous considerations, we set $\alpha = 1.8, 1.85, 1.9$ and

1.95. The diagonal entries of the matrix Σ are $\Sigma_{1,1} = 1.2, \Sigma_{22} = 1.13$ and the off-diagonal is $\Sigma_{1,2} = \Sigma_{2,1} = 0.45$. In the estimation routine, the FFT tuning parameters are initially set to $N = 2^{11}$ and $h = 0.01$, so that the grid, whose endpoints are $\pm Nh/2$, covers [-20.48, 20.48]. In the case there are (centered and scaled) observations outside this initial grid, we increased the distance h to contain all the datapoints for the interpolation procedure.

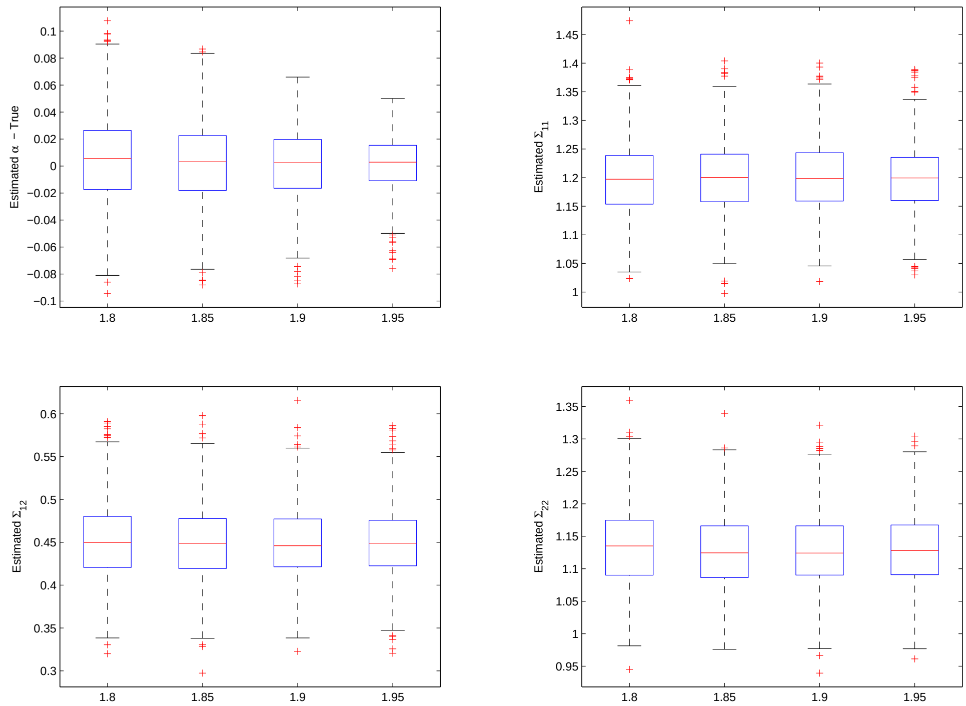


Figure 1.2: Estimated parameter from 1000 simulated $S_\alpha(\Sigma, 0, 0)$ processes for $\alpha = 1.8, 1.85, 1.9, 1.95$ and $\Sigma =$ [1.2 0.45; 0.45 1.13]. Boxes have lines at the 25%, 50% and 75% quantiles.

The simulation results are summarized in form of box plots in Figure 1.2. For the stability index (upper left panel) the box plots indicate that the estimates are centered on the true value (although slightly upward biased) and the dispersion around the median is increasing for lower values of α, along with an increasing left skewness. This is certainly due to the fact that the values α can assume are bounded by 2. The three other panels report the results for the entries of Σ. In this case, irrespective of the values if α, the boxes are centered around the true value and display the same dispersion. This study demonstrates that the estimation procedure, based on the extension of the univariate FFT-ML estimation, delivers reliable estimates of the true parameters of a multivariate Stable distribution.

1.5.3 Estimation results

The estimated parameters for the unconditional model, along with the standard errors, are reported in Table 1.6, second column. Figure 1.3 shows the contour plot of the two series when a

Table 1.6: PG - MRK multivariate DCC-Q-GARCH(1,1) results. Standard errors are reported in parenthesis.

Parameters	Stable	Stable Q-GARCH(1,1)
$\widehat{\alpha}$	1.7191	1.8180
	(0.0180)	(0.0151)
		[1.84]
$\widehat{\sigma_1}$	0.7676	
	(0.0279)	
$\widehat{\sigma_{1,2}}$	0.3295	
	(0.0189)	
$\widehat{\sigma_2}$	1.0447	
	(0.0175)	
$\widehat{a}$	-	0.0250
		(0.0027)
		[0.01]
$\widehat{b}$	-	0.9646
		(0.0063)
		[0.97]
LRT(S-N)	73.4	

For each parameter we report the estimates computed using the one-step approach and below in parenthesis the corresponding standard error. We also report, in square brackets, the estimates of the two-steps approach. LRT(S-N) are the likelihood ratio test statistics of the stable assumption vs. the normal and the Student's t distribution respectively.

multivariate stable model is fitted to the distribution of the returns. The estimate of the stability index α is 1.71 and indicates that the tails of the distributions are heavier than the ones of a normal. An evaluation of the log-likelihood functions favors the stable model. The value of the standard likelihood-ratio test statistic for normal versus stable model, given by

$$LR_{N,S} = -2(Loglik_{normal} - Loglik_{stable}) = 73.4,$$

exceeds the 99%-critical value of the χ_1^2 distribution, which is 6.635. This means a clear rejection of the Gaussian hypothesis. The entries of Σ have no direct interpretation as they are only informative on the covariance matrix of the Gaussian vectors underlying the sub-Gaussian process. One could retrieve, starting from the entries of $\hat{\Sigma}$, the entries of the covariation matrix of the Stable vectors. However, we already mentioned that the covariation matrix, beside being not symmetric, has not the direct interpretation the variance-covariance matrix would have, for example, if a multivariate Gaussian or Student's t distribution were used instead.

For the conditional distribution we decided to use, amongst the numerous multivariate GARCH specifications proposed in literature (see for example Bauwens et al. (2006) for an overview), the dynamic conditional correlation model of Engle (2002) and Tse and Tsui (2002). Our choice follows the one in Bauwens and Laurent (2005), where the same specification is applied to the multivariate skewed t.

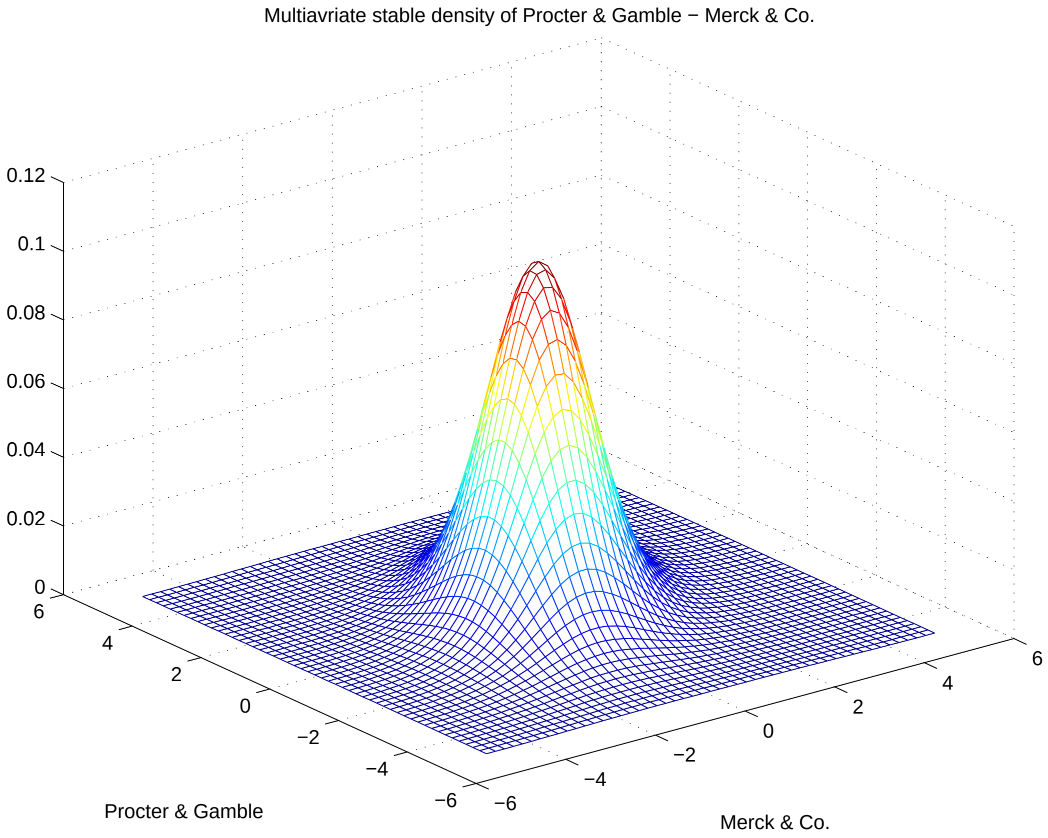


Figure 1.3: Contour plot for Procter & Gamble - Merck & Co.

The dynamic conditional correlation (DCC) model of Engle (2002) is defined as follows:

$$r_t = \mu_t + \Sigma_t^{1/2} z_t, \tag{1.26}$$

$$\Sigma_t = D_t R_t D_t, \tag{1.27}$$

$$D_t = \mathrm{diag}(\sigma_{1,t}, \ldots, \sigma_{k,t}), \tag{1.28}$$

$$R_t = \mathrm{diag}(Q_t)^{-1/2}\ Q_t\ \mathrm{diag}(Q_t)^{-1/2}, \tag{1.29}$$

$$\tag{1.30}$$

where $\mu_t = (\mu_{1t}, ..., \mu_{kt})'$ is the vector of conditional means, Σ_t is the variance-covariance matrix, $\sigma^2_{i,t}$ is the conditional variance of the i-th series specified as univariate GARCH-type equation and R_t is the conditional correlation matrix. Q_t is a $k \times k$ positive definite matrix given by

$$Q_t = (1 - a - b)R + a u_{t-1} u'_{t-1} + b Q_{t-1}, \tag{1.31}$$

where $u_t = (u_{1,t}, \ldots, u_{k,t}),\ u_{i,t} = (r_{i,t} - \mu_{i,t})/\sigma_{i,t}$. R is the $k \times k$ unconditional covariance matrix of the the residuals u_t and a and b are positive scalars. The process is mean reverting as long as $a + b < 1$. R is usually replaced by its empirical counterpart, in order to make the estimation simpler.

In the original model of Engle (2002), the vector r_t is assumed to follow a normal distribution with covariance matrix $\Sigma_t = D_t R_t D_t$. Under the stable hypothesis Σ_t is not the covariance matrix of the returns (it does not exist) but it represents the covariance matrix of the Gaussian vectors underlying the sub-Gaussian process. We want to stress that we are not directly modeling the stable random variables, as there they do not posses the equivalent of the variance-covariance matrix of a Gaussian distribution. We chose instead to model the covariance matrix Σ_t of the Gaussian vector underlying the sub-Gaussian system. Σ_t is a handful object if compared with the covariation matrix of a multivariate stable distribution and, of fundamental importance, any portfolio under the sub-Gaussian assumption is stable distributed with scale parameter that is a linear combination of Σ_t (Proposition 1). Thus, imposing a GARCH-type structure on Σ_t introduces heteroskedasticity also in the multivariate stable model.

It is interesting to note that, even if this Gaussian vectors are a sort of latent random variable, i.e. we cannot split $\epsilon_t = \Sigma_t^{1/2} z_t$ and isolate its stable component A_t and the Gaussian component G_t (see Assumption 1), we can recover the entries of D_t and R.

The conditional scale parameter σ_{it} of the i-th observation $\epsilon_{it} \sim S_\alpha(\sigma_{it,0,0})$ is in fact the square root of the conditional variance of the Gaussian random variable G_{it}. Thus we recover the entries of D_t from the univariate Stable Q-GARCH models. For the entries of R we use the same strategy in combination with the so called projection method used in Nolan (2005). As we said before, R is in the original DCC model the sample correlation matrix of the standardized residuals from the univariate GARCH models. Thus the entries of the principal diagonal will be the square of the scale parameter σ_i of u_i, given that sub-Gaussianity holds also for the residuals of the Stable Q-GARCH model. To estimate the non-diagonal entries of R, i.e. the correlations between G_i and G_j, we use the fact that:

$$\sigma_{i,j} = \frac{1}{2}(\varphi(1,1)^2 - \sigma_i^2 - \sigma j^2), \tag{1.32}$$

where σ_i and σ_j are the scale parameters of the standardized residuals u_i and u_j respectively. $\varphi(1,1)$ is the scale parameter of $\langle(1,1),(u_i,u_j)\rangle = u_i + u_j$. See Nolan (2005) for further details on the projection method.

Our choice to use the DCC model is motivated by two reasons. First, in its formulations, it allows to use different types of univariate GARCH specifications to construct the diagonal matrix D_t and to obtain the standardized residuals u_t. Other models, such as BEKK for instance, are not so flexible. Seconds, its estimation requires a lower computational effort. To avoid the dimensionality problems of most multivariate GARCH models, Engle (2002) shows that the DCC model can be estimated consistently using a two step approach. This is done by first computing the N univariate GARCH models, then, once we obtain the standardized residuals and the matrix

D_t, we compute the parameters in the constant correlation part. In fact, under the normality assumption, the loglikelihood can be written as the sum of a volatility part and a correlation part. One disadvantage of the two steps estimation is that the estimates are not fully efficient and that the standard error must be corrected. But, as long as we are interested in Value-at-Risk forecasting (VaR hereafter), this issue is not relevant. This could affect the estimation of the forecasts' standard error, although Ruiz and Pascual (2002) showed that this does not seem to matter much.

As pointed out in Bauwens and Laurent (2005), when using a non normal distribution, the decomposition proposed by Engle (2002) is no longer possible and one should adopt a one step approach. We follow again their line and, to stay in the spirit of the DCC model, we also propose to estimate the N univariate GARCH-type models (to obtain u_t) and then estimate the correlation part together with the parameter α, the stability index.

The estimated parameters are presented in the second column of Table 1.6, standard error are in parenthesis. Both the one- and two-step (square brackets) methods give similar results. The stability index $\hat{\alpha}$ is 1.81, as expected from the estimates of the $\hat{\alpha}$'s in the univariate setting. It is higher than the unconditional fitting as part of the fat-tail components have been captured by the GARCH filter. The value of 0.96 for $\hat{b}$ indicates that the dynamics of Q_t are quite persistent; nonetheless the process remains stationary as $a + b < 1$.

1.6 Value-at-Risk forecast and evaluation

To further investigate the goodness of our model we perform Value-at-Risk prediction for several portfolios composed of the returns of the two assets we considered in our analysis. Value-at-risk for long trading positions and short trading positions were computed in-sample and out-of-sample.

Let us denote by $\mu_{t+1|\mathbb{I}_t}$ and $\Sigma_{t+1|\mathbb{I}_t}$ the one step ahead forecast of μ_t and Σ_t respectively, given the information available up to time t. The one-step-ahead VaR computed at t for long trading positions is $\mu_{t+1|\mathbb{I}_t}\omega' + q_\gamma\sqrt{\omega\Sigma_{t+1|\mathbb{I}_t}\omega'}$, while for short trading positions it reads $\mu_{t+1|\mathbb{I}_t}\omega' + q_{1-\gamma}\sqrt{\omega\Sigma_{t+1|\mathbb{I}_t}\omega'}$. q_γ denotes the left quantile at $\gamma\%$ of the distribution we assume for the portfolio returns; $q_{1-\gamma}$ is the right quantile at $\gamma\%$. As possible distributions, we considered the normal, the Student's t and the sub-Gaussian. For the stable Paretian case, recall that a linear combination of stable distributed random variables with common tail index follows a stable distribution with the same tail index.

As there are no tabulated values for the quantiles of a symmetric stable distributed random variable with stability index α, we can easily compute the values of the quantiles we are interested in by running Monte Carlo simulations. We simulated a sample of 100,000 observations from a symmetric stable distribution and obtained the empirical quantile $\gamma\%$. The simulation was performed using the method of Chambers-Mallow-Stuck (see Chambers et al., 1976). Thus the

VaR for long trading positions become, under the stable assumption,

$$VaR_{t+1,\gamma} = \mu_{t+1|\mathbb{I}_t}\omega' + s^{\alpha}_{\gamma}\sqrt{\omega\Sigma_{t+1|\mathbb{I}_t}\omega'}$$

with s^{α}_{γ} denoting the empirical quantile γ % obtained via simulation of a $S\alpha S$ distributed random variable.

For the out-of-sample one step ahead VaR computation we followed the same technique as done in Giot and Laurent (2003) and Bauwens and Laurent (2005). The first estimation sample is the complete sample for which data is available less the last 2000[2] observations. The multivariate model is then fitted and the predicted one-day-ahead VaR is compared with the observed return. At the i-th iteration, where i goes from 2 to 2000, the estimation sample is augmented to include one more day and the VaRs are forecasted and recorded. Whenever i is a multiple of 50, the model is re-estimated to update the parameters. Thus we assume a 'stability window' of 50 days for our parameters. The procedure is iterated until all observations less one are included in the observation sample.

After the predicted valued of the VaR are compared with the realized returns of the portfolio, the coverage probability for long and short trading positions are computed. To check whether the empirical percentage of VaR violations is close to the theoretical one we adopt the Kupiec LR test (Kupiec, 1995) and report the p-values for the test for long and short trading positions with different portfolios weights. Table **??** and 1.8 present the in sample and one day ahead out-of-sample VaR evaluation, respectively.

Analyzing the in-sample VaR performance of the three models we see that both Stable and Student's t distributions perform overall well. Under the Student's t assumption we reject the null hypothesis about the true failure rate in only one case. We never reject the null hypothesis under the stable assumption. Normal distribution does not perform particularly well and the null hypothesis is rejected in 7 cases. This is somehow expected given the thickness of the tails that a Gaussian distribution is incapable to capture. *Grade* 5% summarizes the performance of each model and it is defined as the percentage p-values above the 5% critical values (both for long and short trading positions). Since we have only one rejection for the Stable model, its *Grade* 5% is 96 %. For the Student's t it is 90 % and for the normal model only 57%.

We turn now to the one step ahead out-of-sample VaR performance evaluations. Again, the Stable model works well in comparison to the normal and Student's t models, resulting in a value of 92 % of the time the null hypothesis is accepted. The normal model performs badly, only 51 % for the *Grade* 5% value while Student's t results in a percentage of 81 %.

1.7 Extension to higher dimensions

[2]Differently from Giot and Laurent (2003) and Bauwens and Laurent (2005) we perform, given the larger data sample, 2000 one-step-ahead out-of-sample VaR forecasts.

Engle (2007)

Although theoretically straightforward to extend to higher dimensions, the model we propose requires a remarkable computational effort with three or more assets. In the bivariate case we need to create a lattice of dimension $N_1 \times N_2$, where $N_1 = N_2 = 2^{11}$ and use the fast Fourier transform to obtain the p.d.f. Then, via two dimensional linear interpolation, we compute the p.d.f. of the data points falling between the lattice values. In a three-dimensional case, we should create a tensor of dimension $2^{11} \times 2^{11} \times 2^{11}$ and so on for higher dimensions. The memory required to create this multi-dimensional object and the complexity of the n-dimensional interpolation at each maximization step cause a notable increase of the computational burden.

To circumvent this lack of feasibility of our model, we tackle the problem from a different angle. In particular, we show how to extend our Stable-DCC model to higher dimensions using techniques recently introduced in literature to model the covariance matrix for vast dimensional time series. We first present the MacGyver estimator of Engle (2007) and then the composite likelihood (CL) estimator of Engle et al. (2008). This last estimator is more desirable as it requires a lower computational complexity and, unlike the MacGyver estimator, analytical consistency is derived.

1.7.1 Engle's method

The DCC model of Engle (2002) represents a great improvement in solving to so called 'curse of dimensionality' typical of most of the multivariate GARCH models proposed in literature. However, the estimation of correlation matrices for very large systems is not totally solved. This is due to three reasons (Engle, 2007). First, for each observation, the evaluation of the log-likelihood function requires inversion of matrices, R_t, which are full $N \times N$ matrices, N being the number of assets to be considered. To maximize the likelihood function, it is necessary to evaluate the log-likelihood for many parameter values and consequently invert a great many $N \times N$ matrices. Convergence is not guaranteed and sometimes it fails or it is sensitive to starting values. Secondly, Engle and Sheppard (2005) show that in correctly specified models with simulated data, there is a downward bias in the DCC parameter a when the number of asset N is large. Thus the correlations are estimated to be smoother and less variable when a large number of assets are considered than when a small number of assets are considered. Thirdly, there might be structure in correlations which is not incorporated in this specification.

To solve these three problems Engle (2007) proposed the so-called MacGyver estimator. It assumes that the selected DCC model is correctly specified between every pair of assets i and j. Hence, the correlation process of the Gaussian vector underlying the stable model is:

$$\begin{aligned} \rho_{ij,t} &= \frac{q_{ij,t}}{\sqrt{q_{ii,t} q_{jj,t}}} \\ q_{ij,t} &= R_{i,j}(1 - a - b) + a u_{i,t-1} u_{j,t-1} + b q_{ij,t-1}. \end{aligned} \tag{1.33}$$

Because the high dimension model is correctly specified, so is the bivariate model. In each case all bivariate pairs are estimated and then simple aggregation procedures such as means or medians are applied. Simulations of time series of sample size of 1,000 observations with 100 replications and dimensions from $N = 3$ to $N = 50$ and bias of the estimated parameters are also presented. The result is that, according to the root mean square errors (RMS), for a, the medians are the best in most experiments and the median of the unrestricted bivariate parameter estimates has smallest RMS error. For the parameter b, the best estimator for each experiment is either the median or the median of the restricted estimator. The restricted MLE reparameterizes the log-likelihood using a logistic functional form so that both parameters must lie in the interval (0,1).

As stated in Engle (2007), in addition to computational simplification and bias reduction, there are other several advantages to this MacGyver method. When there are 50 assets, there are 1225 bivariate pairs. When there are 100 assets, there are 4950 asset pairs. Hence the number of bivariate estimations increases as well. However, since only median of all these estimations is needed, there is little loss if efficiency if some are not run. This opens the possibility of estimating a subset of the bivariate pairs.

A second advantage is that the data sets for each bivariate pair need not be of the same length. This is of particular importance when, for example, examining large asset classes and cross country correlations as there are many assets which are newly issued, merged or associated with short time histories. A potential third advantage, not explored in Engle (2007), is that there may be evidence in these bivariate parameter estimates that the selected DCC model is not correctly specified. It is assumed in fact that the bivariate models would show less dispersion if the model is correctly specified than if it is incorrect.

In our multivariate Stable-GARCH DCC, using bivariate estimations renders a parallelization of this operation feasible. In fact, whereas a N dimensional log-likelihood maximization cannot be distributed to different machines, each two-dimensional maximization can be parallelized.

1.7.2 The composite-likelihood method

The MacGyver estimator of Engle (2007), if on the one hand overcomes the difficulty of inverting the matrix R_t, on the other hand has the difficulty that: (i) it is not clear that the pooled estimators should have equal weights, (ii) it involves $N(N-1)/2$ maximizations, (iii) no property of this estimator are derived, (iv) the resulting estimator may not be in the permissible parameter space.

A novel and fast way of estimating models of time-varying covariances that overcomes an undiagnosed incidental parameter problem which has troubled existing methods when applied to hundreds or even thousand of assets is proposed in Engle et al. (2008). Their approach is to construct a type of composite likelihood and to maximize it to deliver their preferred estimator. The composite likelihood is based on summing up the quasi-likelihood of subsets of assets. This means that, if as subsets we consider each pair of assets, this approach can be easily adapted to our multivariate stable model.

Assume that a given dataset consists of N assets and denote with r_t the $N \times 1$ vector of returns at time $t, t = 1, \ldots, T$. Consider the general case in which

$$E(r_t|\mathbb{I}_{t-1}) = 0, \quad Cov(r_t|\mathbb{I}_{t-1}) = \Sigma_t, \tag{1.34}$$

Consider the data array $Y_t = \{Y_{1t,\ldots,Y_{Kt}}\}$ where Y_{jt} is itself a vector containing small subsets of the data (there is no requirement for the Y_{ij} to have common dimensions)

$$Y_{jt} = S_j r_t'$$

where S_j is a non-stochastic selection matrix. The example that best fits our bivariate stable model and that is also presented in Engle et al. (2008), is where one looks at all unique "pairs" of data

$$\begin{aligned} Y_{1t} &= (r_{1t}, r_{2t})' \\ Y_{2t} &= (r_{1t}, r_{3t})' \\ &= \vdots \\ Y_{Kt} &= (r_{K-1t}, r_{Kt}) \end{aligned}$$

where $K = N(N-1)/2$. Model (1.34) trivially implies

$$E(Y_{jt}|\mathbb{I}_{t-1}) = 0, \quad Cov(Y_{jt}|\mathbb{I}_{t-1}) = \Sigma_{jt} = S_j \Sigma_t S_j'. \tag{1.35}$$

In our sub-Gaussian model, where $Cov(Y_{jt}|\mathbb{I}_{t-1})$ is not defined, this approach implies that

$$\Sigma_{jt} = S_j \Sigma_t S_j'$$

is the covariance matrix of the pair $(G_{jt}, G_{j+1,t})$ of Gaussian vector underlying the Stable system. Then a valid quasi-likelihood can be constructed for ψ off the j-th subset

$$\log L_j(\psi) = \sum_{t=1}^{T} l_{jt}(\psi), \quad l_{jt}(\psi) = \log f(Y_{jt}; \psi)$$

The quasi-likelihood will have information about ψ but more information can be obtained by averaging the same operation over many submodels

$$c_t(\psi) = \frac{1}{K} \sum_{i=1}^{K} \log L_{jt}(\psi).$$

Of course if the $\{Y_{1t}, \ldots, Y_{Kt}\}$ were independent this would be the exact likelihood. Such functions, based on "submodels" or "marginal models", are called *composite likelihoods* (CLs). Summing over the time series we have the sample CL function

$$CL(\psi) = \sum_{t=1}^{T} c_t(\psi).$$

Evaluation of $c_t(\psi)$ costs $O(K)$ calculations. In the case where all distinct pairs are used this means the CL costs $O(N^2)$ calculations. One can also use contiguous pairs $\{r_{jt}, r_{j+1t}\}$, which would be $O(N)$, or an economically motivated like the called "beta CL" introduced in Engle et al. (2008) which is also $O(N)$ and is based on using all pairs involving the market index returns.

The main assumption of this model is that

$$c_t(\psi) = \frac{1}{N} \sum_{j=1}^{N} \log L_{jt}(\theta, \lambda_j),$$

i.e. it is possible to write the CL in terms of the common finite dimensional θ and then a vector of parameters λ_j which is specific to the j-th pair. The interest is in estimating θ and so the λ_j are nuisances. Although commonality of some elements across the λ_j could be exploited, an alternative strategy is to consider the parameters as variation-free (e.g. Engle et al., 1983):

$$(\lambda_1, \lambda_2, \ldots, \lambda_N) \in (\Lambda_1, \Lambda_2, \ldots, \Lambda_N). \tag{1.36}$$

In this particular setting, inference can be carried out for λ_j based solely on $(Y_{j1}, \ldots, Y_{jN})$ and the common structure determined by θ. This approach, in some sense similar to the two-step estimation of the DCC, risks efficiency loss but not bias.

The estimation strategy can be then generically stated as solving

$$\hat{\theta} = \text{argmax}_{\theta} \frac{1}{K} \sum_{t=1}^{T} \sum_{i=1}^{K} \log L_{jt}(\theta, \hat{\lambda}_j) \tag{1.37}$$

where the nuisance parameters vector for the j-th subset $\hat{\lambda}_j$ solves

$$\sum_{t=1}^{T} g_{jt}(\hat{\theta}, \hat{\lambda}_j) = 0.$$

Here g_{jt} is a $\dim(\lambda_j)$-dimensional moment constraints so that for each j and θ there exists a single $\lambda_{j\theta}$ which solves

$$E(g_{jt}(\theta, \lambda_{j\theta})) = 0, t = 1, 2, \ldots, T.$$

To estimate the parameter of our multivariate Stable-GARCH model using the same DCC specification as before for the matrix Σ_t, we need to estimated for all the $N(N-1)/2$ pairs (or the $N-1$ contiguous pairs), the log-likelihoods $l_{jt}(\theta, \hat{\lambda}_j)$

$$l_{jt}(\theta, \hat{\lambda}_j) = -\frac{1}{2}\log(|\Sigma_{jt}|) + \log f(z_{jt}; \theta) \tag{1.38}$$

where $\theta = (a, b)'$ and $\lambda_j = (\eta_{1j}, \eta_{2j}, \sigma_{1j}, \sigma_{2j}, \sigma_{12j})$. This results in a two-step estimation. In the first step we estimate (η_{1j}, η_{2j}), the parameters of the univariate Stable Q-GARCH and $(\sigma_{1j}, \sigma_{2j}, \sigma_{12j})$, the entries of the correlation matrix R_j of the pair of Gaussian vector underlying the sub-Gaussian process. The entries of R_j are recovered using univariate estimations and the projection method as previously described. In the second step we find $\hat{\theta}$ that satisfies (1.37). This resulting $\hat{\theta}$ is called a m-profile CL estimator (MMCLE).

1.7.3 Illustrative application

In this subsection we demonstrate the feasibility of the model when applied to a relatively large number of stocks. We used the components of the Dow Jones Industrial Average index during the same period of the previous bivariate analysis, i.e. from January 2, 1990 until May 7, 2007. Data were downloaded from *finance.yahoo.com*. Selecting only the companies that have returns throughout the sample (all except KRAFT FOODS INC.) we are left with 29 series of 4372 observations. As done in the previous application of the model, we fitted a stable Q-GARCH(1,1) to model the conditional distribution the series. The estimates of α and β, along with the summability test $\tau_T(\alpha)$ are reported in Table 1.9. The values of α lie between a minimum of 1.79 for 3M COMPANY to a maximum of 1.93 for CHEVRON CO. These relatively high values for the characteristic exponent are somehow expected as some of the fat-tail features of the series has been captured by the Q-GARCH filter. In the fourth columns are reported the estimates of the asymmetry parameter β. The null hypothesis of $\beta = 0$ is rejected at level 95% in only eight series out of 29. This indicates that asymmetry is present in some components but it is not a shared feature among the assets. The last column reports the summability test. At the 90% the null of stable distributed residuals is rejected in 13 cases; at level 95% is rejected in 4 cases and at level 99% is never reject. We can then conclude we do not have elements to fully reject a multivariate stable model for the joint distribution of the 29 components of the Dow Jones index.

In the next step we fitted a DCC models for the entire set of assets using the MacGyver estimator with all the possible pairs and the maximum m-profiled composite-likelihood method with contiguous pairs and all pairs. Again, a stable Q-GARCH filter was adopted to fit each marginal. Results are reported in Table 1.7. The estimated parameters display similar values across the different models. The standard errors when the CL with all pairs is fitted are much higher when compared to the ones delivered by the CL model with contiguous pairs. This loss in efficiency might be due to an imprecise computation of the numerical Hessian matrix given the large number of pairs used in the estimation. This issue requires a further investigation and a

Table 1.7: Result for fitting the DCC model using the MacGyver estimator and maximum m-profiled composite-likelihood (contiguous and all pairs). Estimates for the MacGyver method are obtained taking the median of the 406 bivariate estimates. Numerical standard errors for the composite likelihood estimator are reported in parenthesis.

	MacGyver	m-profile CL	
	(all pairs)	(contiguous pairs)	(all pairs)
α	1.8555	1.8989	1.8984
		(0.0692)	(0.2639)
a	0.0079	0.0120	0.122
		(0.0027)	(0.0584)
b	0.9863	0.9739	0.9743
		(0.0390)	(0.1617)

more comprehensive analysis is left for future research.

As pointed out in Lombardi and Veredas (2009), the application we have presented here is certainly subject to two critiques. First, it is unlikely that the stability index α is the same across all the 29 stocks. Results in Table 1.9 confirm this. Second, some series exhibit a clear deviation from symmetry and the proposed model is restricted to be symmetric. Despite all these shortcomings, this estimation exercise is purely illustrative and proved the feasibility of the multivariate stable-GARCH model even when applied to a high-dimensional system.

1.8 Conclusion

In this paper we proposed a multivariate stable model for the distribution of the asset returns of a portfolio. The class of stable distributions possesses a parameter, the stability index α, which determines the thickness of the tails and thus allows to take into account the phenomenon of excess of kurtosis. Under the hypothesis that the asset returns follow a sub-Gaussian distribution, the joint characteristic function possesses a tractable expressions and this allowed us to estimate the parameters via the likelihood function maximization. The joint density function was recovered by extending to the bivariate case the method of Mittnik et al. (1999a). Moreover, given the particular expression of the characteristic function, we can introduce a multivariate GARCH model for the covariance matrix of the Gaussian vectors underlying the sub-Gaussian system. The scale parameter of the α-stable distributed portfolio returns is, in fact, a linear combination of this covariance matrix. In this way we take into account an other feature typical of financial time series: the heteroskedasticity.

An application to two daily returns for Procter & Gamble and the Merck and Co. stock is presented. The stable model performs better in terms of in-sample and out-of-sample VaR performances when compared with the normal and the Student's t model.

The use of multivariate stable distribution is still a challenge. Our formulation gave us a rather simple expression for the ch.f., but still the computational burden is notable. To overcome this problem, we presented an extension of the model to higher dimensions that follows the recent developments in the vast dimensional time-varying covariances models. In this direction, a deeper analysis needs to be addressed.

Together with volatility clustering and excess of kurtosis, asymmetry is an other important feature present in financial series. Our model is unfortunately restricted to be symmetric. Developing a multivariate stable model which has a tractable expression and allows for asymmetry is without any doubt a challenging task and is left for future research.

Appendix 1.A Proof of Proposition 1

Consider the random variable Y_t,

$$Y_t = \sum_{i=1}^{N} \omega_i X_{it}.$$

If Assumption 1 holds we can write

$$Y_t = A^{1/2} \sum_{i=1}^{N} \omega_i G_{it}.$$

Using the well known property of the Gaussian random variables

$$\sum_{i=1}^{N} \omega_i G_{it} | \mathbb{I}_{t-1} \sim N(0, \omega' \Sigma_t \omega)$$

or, with different notation

$$\sum_{i=1}^{N} \omega_i G_{it} | \mathbb{I}_{t-1} \sim S_2(\frac{\sigma_t}{\sqrt{2}}, 0, 0), \quad \sigma_t^2 = \omega' \Sigma_t \omega.$$

Then, from the definition of sub-Gaussian random variable, it follows that

$$Y_t | \mathbb{I}_{t-1} \sim S_\alpha(\sigma_t, 0, 0)$$

and given that $P_t = Y_t + \omega' \mu$, using the properties of the α-stable distribution

$$P_t | \mathbb{I}_{t-1} \sim S_\alpha(\sigma_t, 0, \omega' \mu).$$

Appendix 1.B Computation of the summability test

We briefly outline here the computation of the stable tail index estimator and the summability test used. For a more detailed account see Paolella (1997), Mittnik and Paolella (1999) and Paolella (2001). The tail index estimator $\hat{\alpha}_{Hint}$ is specifically designed for stable Paretian data and given by:

$$\hat{\alpha}_{Hint} = -0.8110 - 0.3076\hat{\beta} + 2.0278\hat{\beta}^{0.5}, \tag{1.B.1}$$

where $\hat{\beta}$ is the intercept in the sample linear regression of $\hat{\alpha}_{Hill(\mathbf{k})}$ on $\mathbf{k}/1000$; the elements of vector $\mathbf{k}$ are such that $0.2T \leq k \leq 0.8T$ in the steps of max $\{[T/100, 1]\}$, and $\hat{\alpha}_{Hill}(\mathbf{k})$ is the popular Hill (1975) estimator:

$$\hat{\alpha}_{Hill}^{-1} = k^{-1} \sum_{j=1}^{k} \ln\left(Z_{T+1-j:T}\right) - \ln Z_{n-k:T} \tag{1.B.2}$$

with $Z_{j:T}$ denoting the j-th order statistics of the sample $Z_1, \ldots, Z_T$. An accurate approximation to its standard error is given by:

$$\widehat{SE}(\hat{\alpha}_{Hint}) \approx 0.0322 - 0.00205 T_*^{-1} - 0.0008352 T_*^{-2}. \tag{1.B.3}$$

Estimator $\hat{\alpha}_{Hint}$ is unbiased and virtually exact normally distributed in samples as small as 50, with estimated small-sample variance both lower and more accurate than that given in McCulloch (1986).

Consider condensing the T-length vector of (presumably i.i.d. stable Paretian) realizations into summed non-overlapping S-length segments. Let $\hat{\alpha}(s)$ denote the vector of tail-index estimates using the tail estimator $\hat{\alpha}_{Hint}$ evaluated at each element in $s = [1, 2, \ldots, S_{\max}(T)]$, with $S_{\max}(T) = \min(10, [T/200])$. The test statistic is given by:

$$\tau_T = (\alpha) = \hat{b}/\widehat{SE}(\hat{b}), \tag{1.B.4}$$

where $\hat{b}$ denotes the weighted least-squares estimate of the slope of $\hat{\alpha}(s)$ regressed on s (and a constant) and $\widehat{SE}(\hat{b})$, its corresponding standard error. The weights are taken to be the inverse of the estimated standard errors of $\hat{\alpha}(s)$ as given in (1.B.3).

Under the null hypothesis of i.i.d. stable Paretian realizations, we have $\tau_T(\alpha) = 0$ while, under the alternative, $\tau_T(\alpha) > 0$. The sampling distribution of $\tau_T(\alpha)$ is non-standard; simulated γ-level cutoff values under the null hypothesis are approximated by

$$C_T(\alpha, \gamma) = c_{0,\gamma} + c_{1,\gamma}\alpha + c_{2,\gamma}\alpha^2 \tag{1.B.5}$$

for $1 < \alpha < 2$ and $1500 \leq T \leq 10,000$. In practice, an estimated valued of α (obtained using the whole sample) is required. For $\gamma = 0.90, 0.95$ and 0.99, the required coefficients in Eq. (1.B.5) can be expressed as functions of T:

$$\begin{aligned}
C_{0,0.90} &= 6.891 - 0.22591 T_* + 0.1180 T_*^{1/2} \\
C_{1,0.90} &= -3.405 + 0.08045 T_* + 1.479 T_*^{1/2} \\
C_{2,0.90} &= 0.4986 + 0.05480 T_* - 0.8680 T_*^{1/2} \\
C_{0,0.95} &= 8.377 - 0.1007 T_* - 0.0511 T_*^{1/2} \\
C_{1,0.95} &= -1.408 + 0.3862 T_* + 0.0431 T_*^{1/2} \\
C_{2,0.95} &= -1.142 - 0.2334 T_* + 0.3542 T_*^{1/2} \\
C_{0,0.99} &= 14.28 - 0.02562 T_* + 0.2462 T_*^{1/2} \\
C_{1,0.99} &= -4.081 + 0.02769 T_* + 0.6514 T_*^{1/2} \\
C_{2,0.99} &= -0.5839 - 0.005483 T_* - 0.3507 T_*^{1/2}
\end{aligned}$$

where $T_* = T/1000$.

Table 1.8: VaR results for different portfolios of Procter & Gamble - Merck (out-of-sample) .

		Long positions			Short positions		
ω	γ	5%	2.5%	1%	5%	2.5%	1%
	Normal	**0.014**	0.466	0.135	**0.002**	**0.035**	0.093
(1/2 ; 1/2)	Student	0.349	0.466	1.000	0.349	0.051	**0.027**
	Stable	0.406	0.773	0.279	0.297	0.562	0.093
	Normal	**0.007**	0.051	0.510	**0**	0.305	0.821
(7/10 ; 3/10)	Student	0.115	0.305	0.489	0.073	0.101	0.240
	Stable	0.468	0.380	1.000	0.297	0.562	1.000
	Normal	**0.034**	0.562	0.058	**0**	**0**	0.052
(3/10 ; 7/10)	Student	0.349	0.664	0.510	**0.014**	**0.023**	**0.027**
	Stable	0.468	1.000	0.279	0.115	0.466	**0.027**
	Normal	0.115	0.886	**0.013**	**0.019**	**0.073**	0.383
(-4/10 ; 14/10)	Student	0.115	0.562	0.510	0.115	0.138	0.154
	Stable	0.209	0.886	0.383	0.115	0.773	0.352
	Normal	**0**	**0.010**	0.658	**0.010**	0.305	**0.036**
(14/10 ; -4/10)	Student	**0.002**	0.073	0.489	0.115	0.239	0.383
	Stable	**0.004**	0.101	0.510	0.534	0.886	**0.036**
	Normal	**0.044**	0.562	**0.090**	**0.001**	**0.002**	0.240
(2/10 ; 8/10)	Student	0.142	0.562	0.383	**0.034**	0.051	**0.013**
	Stable	0.468	0.776	0.279	0.073	0.184	0.093
	Normal	**0.007**	0.101	1.000	**0.014**	0.466	0.821
(8/10 ; 2/10)	Student	0.073	**0.015**	0.489	0.057	0.466	0.093
	Stable	0.115	0.239	0.821	0.605	0.466	0.489
	Normal	**0.034**	1.000	0.058	**0.003**	**0.002**	0.489
(1/10 ; 9/10)	Student	0.073	0.664	0.383	**0.044**	0.051	**0.027**
	Stable	0.297	0.773	0.279	0.173	0.239	0.154
	Normal	**0.003**	0.184	0.658	**0.005**	0.138	0.510
(9/10 ; 1/10)	Student	**0.010**	0.138	0.489	**0.034**	0.138	0.240
	Stable	**0.034**	0.380	0.824	0.115	0.305	0.648
		Grade (5 %)					
	Normal	51 %					
	Student	81 %					
	Stable	92 %					

The entries are the p-values of the null hypothesis: $\hat{f}_l = \gamma$ and $\hat{f}_s = \gamma$, where $\hat{f}_l$ and $\hat{f}_s$ are the failure rate for long and short position respectively. ω is the vector of weights of the portfolio. *Grades* 5 % reports for the three models the percentage of p-values above the 5 % critical value, both for long and short positions.

Table 1.9: Selected parameters estimates and summability inference measures for the components of the Dow Jones Industrial Average index. The period goes from January 2, 1990 to May 7, 2007.

Ticker	$\hat{\alpha}$	$\widehat{SE}(\alpha)$	$\hat{\beta}$	$\widehat{SE}(\beta)$	$\hat{V}_Q$	$\widehat{SE}(V_Q)$	$\tau_T(\alpha)$
AA	1.87	(0.0184)	$0.270^{\dagger}$	(0.1012)	0.993	(0.0015)	2.63^{*}
AXP	1.88	(0.0179)	$0.290^{\dagger}$	(0.1105)	0.993	(0.0008)	2.42
BA	1.83	(0.0190)	0.114	(0.0837)	0.994	(0.0036)	3.05^{*}
BAC	1.9	(0.0167)	-0.187	(0.1346)	0.994	(0.0007)	-0.26
C	1.88	(0.0173)	0.135	(0.1125)	0.996	(0.0017)	1.01
CAT	1.83	(0.0199)	$0.215^{\dagger}$	(0.0807)	0.993	(0.0008)	1.46
CVX	1.94	(0.0131)	-0.260	(0.1869)	0.976	(0.0036)	2.61^{*}
DD	1.89	(0.0173)	0.181	(0.1117)	0.996	(0.0016)	2.67^{*}
DIS	1.86	(0.0187)	0.161	(0.0966)	0.995	(0.0018)	2.43
GE	1.92	(0.0135)	0.011	(0.1629)	0.992	(0.0019)	1.91
GM	1.84	(0.0213)	$0.209^{\dagger}$	(0.0861)	0.988	(0.0027)	2.59^{*}
HD	1.88	(0.0172)	0.078	(0.1104)	0.988	(0.0023)	2.28
HPQ	1.85	(0.0186)	0.089	(0.0905)	0.998	(0.0019)	1.49
IBM	1.81	(0.0197)	0.137	(0.0754)	0.991	(0.0025)	4.6^{**}
INTC	1.87	(0.0168)	-0.055	(0.1064)	0.993	(0.0024)	2.97^{*}
JNJ	1.88	(0.0174)	0.169	(0.1051)	0.992	(0.0015)	4.39^{**}
JPM	1.87	(0.0186)	0.033	(0.1065)	0.996	(0.0023)	2.70^{*}
KO	1.85	(0.0179)	$0.212^{\dagger}$	(0.0910)	0.998	(0.0013)	1.25
MCD	1.87	(0.0172)	$0.219^{\dagger}$	(0.0984)	0.994	(0.0025)	2.98^{*}
MMM	1.8	(0.0210)	$0.160^{\dagger}$	(0.0703)	0.994	(0.0021)	4.52^{**}
MRK	1.85	(0.0196)	-0.016	(0.0946)	0.985	(0.0028)	1.64
MSFT	1.85	(0.0186)	0.238	(0.0912)	0.996	(0.0027)	2.65^{*}
PFE	1.89	(0.0173)	-0.052	(0.1293)	0.990	(0.0030)	0.55
PG	1.86	(0.0183)	0.075	(0.009)	0.993	(0.0049)	1.13
T	1.89	(0.0170)	0.110	(0.1190)	0.994	(0.0021)	4.43^{**}
UTX	1.87	(0.0203)	0.111	(0.1062)	0.990	(0.0025)	0.41
VZ	1.9	(0.0159)	0.205	(0.1212)	0.992	(0.0024)	0.73
WMT	1.9	(0.0163)	$0.272^{\dagger}$	(0.1190)	0.997	(0.0014)	3.05^{*}
XOM	1.93	(0.0134)	-0.067	(0.1635)	0.991	(0.0024)	-0.86

Column $\tau_T(\alpha)$ is the summability test statistics; no stars means we cannot reject the null of stability at the 90% (and then also the 95% and 99%).
*The null of stability at 95% level cannot be rejected here.
**The null of stability at 99% level cannot be rejected here.
†Significant at 95% level.

Part II

Realized volatility models

Manuscript 2

Forecasting realized (co)variances with a block structure Wishart autoregressive model

[1]

[1] This manuscript is a joint work with Angelo Ranaldo (Swiss National Bank) and Massimiliano Caporin (University of Padua)-The views expressed herein are those of the authors and not necessarily those of the Swiss National Bank, which does not accept any responsibility for the contents and opinions expressed in this paper-

2.1 Introduction

The increased availability of high-frequency data provides new tools for forecasting variances and covariances between assets. In particular, after the seminal paper by Andersen and Bollerslev (1998), the literature on realized volatility has grown enormously; see McAleer and Medeiros (2006) for a review.

While most works focus on the study of univariate series, recently there has been growing theoretical and empirical interest in extending the results for the univariate process to a multivariate framework. In this context, two pioneering contributions have been made by Barndorff-Nielsen and Shephard (2004) and Bandi and Russel (2005). Barndorff-Nielsen and Shephard (2004) did not consider the presence of microstructure noise, whereas of the noise has been considered in Bandi and Russel (2005).

Alternative approaches to the high-frequency covariance estimator have recently been introduced by Hayashi and Yoshida (2005, 2006), Sheppard (2006) and Zhang (2006), among others. For example, instead of using calendar returns, the Hayashi and Yoshida estimator (HY) is based on overlapping tick-by-tick returns. Sheppard (2006) analyzed the conditions under which the realized covariance is an unbiased and consistent estimator of the integrated covariance. Zhang (2006) also studied the effects of microstructure noise and non-synchronous trading in the estimation of integrated covariance between assets.

Although the literature on multivariate extensions of the realized variance regarding the definition of new estimators of the realized covariances resulted in a notable amount of academic works, only a few papers provide financial applications for these new estimators.

One explanation for the scarcity of empirical contributions in multivariate realized volatility analysis is the difficulty in finding a dynamic specification of a stochastic volatility matrix which satisfies the symmetry and positivity properties of each forecasted matrix, does not suffer from the so called 'curse of dimensionality' and possesses a closed-form expression for the forecasts at any horizon.

In an interesting paper, de Pooter et al. (2006) investigate the benefits of high-frequency intraday data when constructing mean-variance efficient stock portfolios with daily rebalancing from the individual constituents of the S&P 100 index. The author analyzed the issue of determining the optimal sampling frequency, as judged by the performances of the estimated portfolios. As in Fleming et al. (2001, 2003), and building on the work of Foster and Nelson (1996) and Andreou and Ghysels (2002), in this paper a rolling window volatility estimator is used to forecast the conditional variance matrix $V_{t,h}$:

$$\widehat{V}_{t,h} = \exp(-\alpha_h)\widehat{V}_{t-1,h} + \alpha_h \exp(-\alpha_h) Y_{t-1} \tag{2.1.1}$$

where α_h can be estimated by means of maximum likelihood for the model

$$r_t = \widehat{V}_{t,h}^{1/2} z_t \tag{2.1.2}$$

with $z_t \overset{i.i.d.}{\sim} N(0, I)$ and Y_t as the realized covariance matrix estimated using I intraday returns of equal length $h \equiv 1/I$. r_t is the usual $n \times 1$ vector of daily returns at time t of the n assets composing the portfolio.

In a related paper, Bandi et al. (2006) evaluate the economic benefits of methods that have been suggested to optimally sample (in a MSE sense) high-frequency returns data for the purpose of realized variance and covariance estimation in the presence of market microstructure noise. However, their approach is different from that in de Pooter et al. (2006); their method is designed to select the time-varying optimal sampling frequency for each entry in the covariance matrix based on MSE criteria. Subsequently, the economic gains yielded by the MSE-based optimal sampling are evaluated by comparing the utility gains provided by optimally sampled realized covariance with realized covariances based on fixed intervals. To forecast each entry of the covariance matrix, they adopted an ARFIMA$(2, d, 2)$ model.

An alternative way to forecast the realized variance/covariance matrix is to adopt a matrix transformation that guarantees the positive definitiveness of the forecasts.

Bauer and Vorkink (2007) present a new matrix logarithm model of realized covariance stock returns which uses latent factors as functions of both lagged volatility and returns. The model has several advantages in that it is parsimonious, does not impose parametric restrictions, and yields positive definite covariance matrices.

In Chiriac and Voev (2008) a model based on a multivariate, fractionally integrated autoregressive moving average (ARFIMA) process for the elements of the Cholesky factors of the observed matrix series is proposed. Denoting with Y_t the $n \times n$ realized covariance matrix at time t, with n the number of assets considered, the Cholesky decomposition of Y_t is given by the upper triangular matrix P_t, for which $P_t P_t' = Y_t$. Then the following model is used

$$\Phi(L) D(L)(X_t - \mu) = \Theta(L)\epsilon_t, \qquad \epsilon_t \sim N(0, \Sigma_t). \tag{2.1.3}$$

$X_t = vech(P_t)$ is the vector obtained by stacking the upper triangular components of the matrix P_t in a vector, $\Phi(L)$ and $\Theta(L)$ are matrix lag polynomials and $D(L) = diag[(1-L)^{d_1}, \ldots, (1-L)^{d_m}]$, where $d_1, \ldots, d_m$ are the degrees of fractional integration of each of the m elements of the vector X_t. μ is a vector of constants. Parameters in (2.1.3) are not directly interpretable. However, the dynamic linkages among the variances and covariances series as functions of those parameters are derived.

While both the matrix logarithmic transformation and the Cholesky decomposition have the advantage of guaranteeing the positive definiteness of the covariance matrix, they also have a major drawback: the coefficients of the model totally rule out any possible interpretation. In other words, there is no way to check the significance of the interactions between variances and covariances and thus to reduce the number of parameters in the model by imposing no or limited spillover between the variances and covariances.

A solution to this problem is represented by the Wishart autoregressive model (WAR) proposed by Gourieroux et al. (2009). The model is based on a dynamic extension of the Wishart distribution. This specification is compatible with financial theory, satisfies the constraints on volatility matrices, has a flexible form and, most importantly, maintains the coefficients' interpretability.

The main innovation proposed in this paper is the introduction of a specific parametrization of the WAR model. In particular, we show how to achieve a great reduction of the number of parameters according to an economic criterion which is consistent with standard sectorial asset allocation approaches. The parametric structure we propose imposes a block structure on the coefficient matrices, hence we name the model *block WAR*. The use of block structures in parameter matrices is similar to that in Billio et al. (2006), Billio and Caporin (2008), Asai et al. (2008). Engle and Kelly (2008) introduce a block structure for the correlation matrix while Caporin and Paruolo (2008) present a spatial solutions to the course of dimensionality problem in multivariate volatility models that implies a block structure on the coefficient matrices. In this paper we assume that the asset variances-covariances have no or limited spillover and that their dynamic is sector-specific. A pairwise preliminary analysis confirms this assumption and allows us to substantially reduce the number of parameters implied by the model. In addition, we propose a Wishart-based generalization of the HAR model of Corsi (2009), named HAR-WAR model. We present an empirical application based on variance forecasting and risk evaluation of a portfolio of two US treasury bills (T-bills) and two exchange rates. We compare our restricted specifications with the traditional WAR parameterizations. Our results show that the restrictions may be supported by the data and that the risk evaluations of the models are extremely close. This confirms that our model can be safely used in a large cross-sectional dimension given that it provides results similar to fully parameterized specifications.

In modeling and forecasting volatility, two main trade-offs emerge: mathematical tractability at detriment of economic interpretation and being precise or fast. Our model is an attempt to reconcile, at least partially, both trade-offs. The former trade-off is crucial for many financial applications, including portfolio and risk management. The speed-accuracy trade-off is more and more relevant if we consider the burgeoning phenomenon of algorithmic trading[2].

Section 2.2 introduces the WAR model of Gourieroux et al. (2009), followed by our proposed generalization. Section 2.3 presents the estimation procedure and show an alternative way to estimate the degrees of freedom of the model, a key element to determine if the density of the Wishart distribution exists. The dataset we used is presented in Section 2.4 and an empirical

[2]For instance, using a unique database provided by the Electronic Broking Services (EBS) Chaboud et al. (2009) show that the participation rate of algorithmic trading to the EUR/USD and USD/CHF turnover in 2008 was more than 50% (80%).

application based on portfolio risk evaluation is provided in Section 2.5. Section 2.6 concludes and gives directions for future research.

2.2 The block Wishart autoregressive model

In the following we define the basic Wishart auto regressive model of Gourieroux et al. (2009) and then we introduce the set alternative parametric restrictions that define the block WAR.

2.2.1 The Wishart autoregressive process

Denote by Y_t the time t (realized) covariance for a group of n assets. The sequence of stochastic positive definite Y_t matrices is said to follow a Wishart process if the following relations hold.

At first, the (realized) covariance may be represented as a sum of underlying stochastic processes

$$Y_t = \sum_{k=1}^{K} x_{k,t} x'_{k,t}, \tag{2.2.1}$$

where $x_{k,t}, k = 1, 2, \ldots, K$ are independent Gaussian VAR(1) processes of dimension n with a common autoregressive parameter matrix M and common innovation variance Σ:

$$x_{k,t} = M x_{k,t-1} + \epsilon_{k,t}, \qquad \epsilon_{k,t} \overset{i.i.d.}{\sim} N(0, \Sigma). \tag{2.2.2}$$

When Y_t is defined as in (3.2.1) and (3.2.2) we say it follows a WAR process of order 1, denoted $W[K, M, \Sigma]$. The transition density of WAR(1) depends on the following parameters: K, the scalar degree of freedom (the number of underlying VAR processes), strictly greater than $n-1$ (the number of assets minus one); M, the $n \times n$ matrix of autoregressive parameters; and Σ, the $n \times n$ symmetric and positive definite matrix of innovation covariances. An important property of the Wishart distribution is that the matrices Y_t are positive definite if and only if $K \geq n$ and for a non-centered Wishart specification, the distribution of Y_t possesses a density function only when $K > n-1$ (hence the condition above). Thus, for $K < n-1$ no density can be defined and for $K < n$ the process Y_t is given by a sequence of singular covariance matrices with degenerate Wishart distribution (Muirhead, 1982). We stress that the interpretation of Y_t from latent Gaussian VAR(1) processes is valid for integer valued K only and, in general, any economic or financial interpretation of the latent processes $(x_{k,t})$ is not necessary. The dynamic of a Wishart autoregressive process for any $K > n-1$ is specified by its conditional Laplace transform, which defines the conditional expectations of any exponential transformation of element of the matrix Y_{t+1} (see Gourieroux et al. (2009) for more details):

$$\begin{aligned} \mathbf{\Psi}_t(\Gamma) &= E[\exp Tr(\Gamma Y_{t+1})] \\ &= \frac{\exp Tr\left[M'\Gamma(\mathbf{I}_d - 2\Sigma\Gamma)^{-1} M Y_t\right]}{[\det(\mathbf{I}_d - 2\Sigma\Gamma)]^{K/2}}. \end{aligned}$$

In this paper we follow the line of Gourieroux et al. (2009), in which the latent processes are introduced mainly to provide an intuitive understanding of parameters and results.

From Proposition 2 in Gourieroux et al. (2009) we have:

$$E_t\left(Y_{t+1}\right) = MY_tM' + K\Sigma. \tag{2.2.3}$$

The first conditional moment is thus an affine function of the lagged values of the volatility process. In particular, the WAR(1) process is a weak linear AR(1) process. More precisely we get:

$$Y_{t+1} = MY_tM' + K\Sigma + \eta_{t+1}, \tag{2.2.4}$$

where η_{t+1} is a matrix of stochastic errors with a zero conditional mean. Equivalently, we may represent Y_t conditional mean in the following companion form:

$$vech(Y_{t+1}) = A(M)vech(Y_t) + vech(K\Sigma) + vech(\eta_{t+1}), \tag{2.2.5}$$

where $vech(Y)$ denotes the vector obtained by stacking the lower triangular elements of Y, and $A(M)$ is a function of M. The error term η is a weak white noise, since it features conditional heteroskedasticity and, even after conditional standardization, is not identically distributed.

In general, WAR processes with higher autoregressive order p may be considered and the Wishart process can be easily extended to include more autoregressive lags. This is accomplished by replacing the conditioning matrix MY_tM' with any symmetric positive semi-definite function of $Y_t, Y_{t-1}, \ldots, Y_{t-p+1}$. However, when the autoregressive order is larger than 1, the interpretation of the Wishart process as the sum of squares of autoregressive Gaussian processes is no longer valid even for integer K. For a WAR(p) process, the equivalent of (3.2.4) reads:

$$E_t\left(Y_{t+1}\right) = \sum_{j=1}^{p} MjY_{t+1-j}M_j' + K\Sigma. \tag{2.2.6}$$

In the following, unless differently stated, we will refer only to WAR(1) specifications.

2.2.2 Interpretation of the coefficients

The principal drawback of many multivariate volatility models is the so-called 'curse of dimensionality', that is, the number of parameters is a power function of the cross-sectional model dimension. One of the main contributions of this paper is to provide a sensible reduction of the parameter space by imposing a set of restrictions on the standard WAR model. Our modeling approach will be presented in the following section; here we provide the intuition on parameter interpretation within the WAR model.

In the simple case of a (2×2) matrix, as done in Gourieroux (2007), we define the best prediction of Y_t given by a WAR(1) model. Then we present the approaches we suggest to reduce the parameter space.

Consider the (2×2) covariance matrix Y_t, the autoregressive matrix M and the innovation variance Σ:

$$Y_t = \begin{pmatrix} Y_{11,t} & Y_{12,t} \\ Y_{12,t} & Y_{22,t} \end{pmatrix}, M = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$$

The full WAR(1) model specifies the best prediction of Y_t, $E[Y_t|Y_{t-1}]$ as:

$$E[Y_t|Y_{t-1}] = \begin{pmatrix} a_1Y_{11,t-1} + b_1Y_{12,t-1} + c_1Y_{22,t-1} + d_1 & a_2Y_{11,t-1} + b_2Y_{12,t-1} + c_2Y_{22,t-1} + d_2 \\ - & a_3Y_{11,t-1} + b_3Y_{12,t-1} + c_3Y_{22,t-1} + d_3 \end{pmatrix} \tag{2.2.7}$$

where a_j, b_j, c_j and $d_j,\ j = 1, \ldots, 3$ are scalar parameters. d_j corresponds to K times the entries of Σ. By construction, the prediction is a symmetric semi-definite positive matrix for any Y_{t-1} which belong to $\mathcal{S}^+$, the set of symmetric positive definite matrices. To express it in terms of M we have:

$$\begin{cases} a_1 = m_{11}^2, & b_1 = 2m_{11}m_{12}, & c_1 = m_{12}^2, \\ a_2 = m_{11}m_{21}, & b_2 = m_{11}m_{22} + m_{21}m_{12}, & c_2 = m_{12}m_{22}, \\ a_3 = m_{21}^2, & b_3 = 2m_{21}m_{22}, & c_3 = m_{22}^2, \end{cases}$$

The effect of the past variances and covariances on the present volatility can be seen immediately. First, note that the full WAR model allows for spillover between variances and covariances.

Therefore, a possible strategy is to reduce the numbers of parameters by assuming no or limited spillover between the variances. For instance, setting $m_{12} = 0$ implies that the conditional variance of the first asset depends only on its past shocks and that the second asset variance does not influence the conditional covariance. Differently, a diagonal specification of M corresponds to the absence of spillovers between variances and covariances.

Those restrictions on the dynamic model are clearly related with non-causality restriction concerning volatilities and covolatilities. Linear (in the Granger sense) and nonlinear causalities are investigated and compared, for a bivariate WAR process, in Jasiak and Lu (2007). Gourieroux and Sufana (2007) characterize nonlinear causality hypothesis for model based on the conditional Laplace transform (the WAR process being one of those) and provide interpretations of the linear and quadratic causality in this framework.

In particular, in the bivariate WAR of order 1, the Granger noncausality relations are defined as

(1) $(Y_{12}, Y_{22})' \nrightarrow Y_{11} \Leftrightarrow E[Y_{11,t+1}|Y_{11,t}, Y_{22,t}, Y_{12,t}] = E[Y_{11,t+1}|Y_{11,t}]$

(2) $(Y_{11}, Y_{12})' \nrightarrow Y_{22} \Leftrightarrow E[Y_{22,t+1}|Y_{11,t}, Y_{22,t}, Y_{12,t}] = E[Y_{22,t+1}|Y_{22,t}]$

(3) $(Y_{11}, Y_{22})' \nrightarrow Y_{12} \Leftrightarrow E[Y_{12,t+1}|Y_{11,t}, Y_{22,t}, Y_{12,t}] = E[Y_{12,t+1}|Y_{12,t}]$

(4) $Y_{11} \nrightarrow (Y_{12}, Y_{22})' \Leftrightarrow E[(Y_{12,t+1}, Y_{22,t+1})'|Y_{11,t}, Y_{22,t}, Y_{12,t}] = E[(Y_{12,t+1}, Y_{22,t+1})'|Y_{12,t}, Y_{22,t}]$

(5) $Y_{12} \nrightarrow (Y_{11}, Y_{22})' \Leftrightarrow E[(Y_{11,t+1}, Y_{22,t+1})'|Y_{11,t}, Y_{22,t}, Y_{12,t}] = E[(Y_{11,t+1}, Y_{22,t+1})'|Y_{11,t}, Y_{22,t}]$

(6) $Y_{22} \nrightarrow (Y_{11}, Y_{12})' \Leftrightarrow E[(Y_{11,t+1}, Y_{12,t+1})'|Y_{11,t}, Y_{22,t}, Y_{12,t}] = E[(Y_{11,t+1}, Y_{12,t+1})'|Y_{11,t}, Y_{12,t}]$

where the symbol $\nrightarrow$ indicates the absence of Granger causality. The sufficient and necessary conditions for Granger linear noncausality are:

(1) $(Y_{12}, Y_{22})' \nrightarrow Y_{11} \Leftrightarrow m_{12} = 0$

(2) $(Y_{11}, Y_{12})' \nrightarrow Y_{22} \Leftrightarrow m_{21} = 0$

(3) $(Y_{11}, Y_{22})' \nrightarrow Y_{12} \Leftrightarrow m_{11}m_{21} = 0$ and $m_{12}m_{22} = 0$

(4) $Y_{11} \nrightarrow (Y_{12}, Y_{22})' \Leftrightarrow m_{21} = 0$

(5) $Y_{12} \nrightarrow (Y_{11}, Y_{22})' \Leftrightarrow m_{11}m_{12} = 0$ and $m_{21}m_{22} = 0$

(6) $Y_{22} \nrightarrow (Y_{11}, Y_{12})' \Leftrightarrow m_{12} = 0$

In the case in which M is diagonal, i.e. when $m_{12} = m_{21} = 0$, all noncausality relations (1)-(6) are satisfied and we have

$$\begin{aligned}
Y_{11,t+1} &= m_{11}^2 Y_{11,t} + K\sigma_{11} + \eta_{11,t+1}, \\
Y_{12,t+1} &= m_{11}m_{22} Y_{12,t} + K\sigma_{12} + \eta_{12,t+1}, \\
Y_{22,t+1} &= m_{22}^2 Y_{22,t} + K\sigma_{22} + \eta_{22,t+1},
\end{aligned}$$

and thus each entry of Y_t depends only on its past values.

This very simple example in two dimensions helps us to identify the coefficients in M that plays a role in the spillover effect between variances. Using the delta method we can, in fact, easily compute the standard errors for the a_i, b_i and c_i and thus evaluate which parameters are significant and check the appropriateness of assumption of limited spillover. We will present now four different parametrizations for the WAR process that impose no or limited spillover. We also show in the empirical analysis that the restrictions we impose on the matrix M are justified by the data.

2.2.3 Specifications of the block Wishart autoregressive model

To derive the block WAR model we impose a set of restrictions on the matrix M. These restrictions come from a criterion allowing assets to be grouped. Some examples are given by the economic sector of the stocks entering into an equity portfolio, the type of assets entering into a diversified equity-bond portfolio, or the geographical reference areas of a group of assets. The main intuition behind asset grouping is that the clustered variables may share common patterns or common features, and that their variance-covariance dynamic is similar. In fact, we can presume that assets belonging to the same economic sector may have a similar reaction to market shocks/news, and are similarly affected by market movements.

Clearly, groups may be defined on a data-driven basis, such as referring to the dynamic properties of the series mean and/or variances, or on mixed criteria. The comparison of alternative methods for clustering financial assets is outside the scope of this paper and will not be considered. In the following we will use *a priori* defined groups in order to present our modeling approach and to show, on an empirical basis, its advantages.

Consider the simple WAR(1) model as in Eq. (3.2.5):

$$Y_{t+1} = MY_tM' + K\Sigma + \eta_{t+1}.$$

Assume that our portfolio consists of n stocks and that we can classify them into N groups, according to some economic (or data-driven) criterion, as discussed in the previous section (such as the economic sector or the existence of common patterns in realized variances and covariances).

The N groups have dimension n_i with $\sum_i n_i = n$. In addition, the assets are ordered following a group rule, that is, assets from 1 to n_1 belong to group 1, assets from n_1+1 to n_1+n_2 belongs to group 2, and so on. Given this asset classification, the autoregressive matrix M may be partitioned as follows:

$$M = \begin{pmatrix} M_{11} & \cdots & M_{1N} \\ \vdots & M_{ii} & \vdots \\ M_{N1} & \cdots & M_{NN} \end{pmatrix},$$

where M_{ij} is a matrix of dimension $n_i \times n_j$.

By imposing a particular structure on the matrices M_{ij} we be able to reduce the number of parameters of the model. We propose the following specifications:

(i) $M_{ij} = \mathbf{0} \quad \forall i \neq j,\ i,j = 1,\ldots,N,$

(ii) $M_{ij} = \mathbf{0}$ and $M_{ii} = \alpha_i(\mathbf{1}_{n_i}\mathbf{1'}_{n_i}), \quad \forall i \neq j,\ i,j = 1,\ldots,N$

(iii) $M_{ij} = \mathbf{0}$ and $M_{ii} = (\alpha_{i,1},\ldots,\alpha^{i,n_i})(\mathbf{I}_{n_i}), \quad \forall i \neq j,\ i,j = 1,\ldots,N$

(iv) $M_{ij} = \mathbf{0}$ and $M_{ii} = \alpha_i(\mathbf{I}_{n_i}), \quad \forall i \neq j,\ i,j = 1,\ldots,N$

where $\mathbf{1}_{n_i}$ is a $n_i \times 1$ vector of ones and $\mathbf{I}_{n_i}$ is the identity matrix of dimension n_i.

If assets belonging to the same group share common reactions to shocks, we can hypothesize, to some extent, that their co-volatilities also have a similar behavior. If the groups are sector-specific, model (i) implies that the variances and covariances of each asset are only influenced by the variances and covariances of assets belonging to the same class. Therefore, no volatility spillover exists between assets belonging to different sectors. We named this model *block WAR.* The number of parameters that needs to be estimated is $n(n+1)/2 + \sum_{i=1}^{N} n_i^2$, along with the degrees of freedom K.

A further reduction of the number of parameters is obtained by imposing a single parameter for each group, as shown in model (ii). In this case, the variance and covariance of each asset belonging to, say, group j depends on the past values of itself, on the past values of the variances of the other assets of the same group and on the covariances with those assets via a function of the unique parameter α_j. We call this model *restricted block WAR.* This models contains $n(n+1)/2 + N$ parameters in M and Σ plus K.

Model (iii) relaxes the assumption of spillover between assets belonging to the same sector. It assumes each matrix $M_{ii},\ i = 1, \dots, N_i$ to be diagonal, i.e. the autoregressive matrix M is diagonal. In this case grouping the assets according to some criterion does not affect the parametric space. We named this model *diagonal WAR.* For this model, n parameters need to be estimated in the matrix M, plus the $n(n+1)/2$ parameters in Σ and the degrees of freedom K. One of the implications of the diagonal structure for M is that each realized variance is only a function of its past values.

If we assume again that assets belonging to the same sector have common dynamics for the variance, or if we can find a way to group assets whose volatilities obey the same process, the number of parameters can be further reduced. This is the case for model (iv). For each group a single parameter is taken to model the dynamics of the variances for the assets in the considered group, i.e. the elements on the diagonal of each $M_{ii},\ i = 1, \dots, N$ are all equal. In total only $N + n(n+1)/2 + 1$ parameters are required in this model. We refer to this model as the *restricted diagonal WAR.*

It is worth mentioning that the specifications (i)-(iv) are only a subset of all the possible specifications of the WAR model. In fact, we set all the off-diagonal blocks to zero. The assumption $M_{ij} = \mathbf{0}\ \forall i \neq j, i, j = 1, \dots, N$ can be replaced by the same structure we imposed on the matrices M_{ii}: full, scalar, diagonal and restricted diagonal. This allows us to consider not only the interactions between assets belonging to the same group, but also interactions between a limited set of groups. Finally, we highlight that block structured WAR representations induce some restrictions on causality across the variances and covariances of asset *groups.* Under (i) we impose noncausality between the variances and covariances of different asset groups. Under (ii) we also include a common structure of causality within asset groups variances and covariances. Moreover, (iii) and (iv) impose noncausality across variances and covariances. In this paper we stick with a structure that ignores the off-diagonal blocks and leave a full generalization of the WAR model for future works.

2.2.4 The block HAR-WAR model

One of the stylized facts about asset returns is the long-run temporal dependencies of return volatilities. The literature on volatility modeling has documented that such temporal dependencies are highly persistent. In particular, the low first-order autocorrelations usually found in empirical analysis (Thomakos and Wang, 2003), along with their slow decay, suggest that the logarithmic realized standard deviations do not contain a unit root but exhibit long memory.

To account for this, fractionally integrated autoregressive models (ARFIMA) have been shown to be effective in empirical modeling (see Andersen et al. (2001a) and Andersen et al. (2001b) among others). Fractional integration achieves long memory parsimoniously by imposing a set of infinite dimensional restrictions on the infinite variable lags but completely lacks a clear mathematical interpretation.

Another crucial point is that the long memory observed in the data could be only an apparent behavior generated from a process which is not really long memory. Indeed, the usual tests can indicate the presence of long memory simply because the largest aggregation level that we are able to consider is not large enough. LeBaron (2001) shows that a very simple additive model defined as the sum of only three different linear processes (AR(1) processes) each operating on a different time frame, can display hyperbolic decaying memory, provided that the longest component has a half-life that is long relative to the test aggregation ranges. Another result from Granger (1980) shows that the sums of an high number of short memory processes can induce long memory. In Pong et al. (2004) an ARMA(2,1) is proposed to model and forecast realized volatility. The authors' choice is motivated by the research of Gallant et al. (1999), who show that the sum of two AR(1) processes is capable of capturing the persistent nature of asset price volatility. In their paper Pong et al. (2004) show that the short memory ARMA(2,1) model is as good as long memory ARFIMA models when forecasting futures volatilities. Motivated by the existence of multiple volatility components in intraday frequencies, along with the apparent long-memory characteristic, Andersen and Bollerslev (1997) formulated a version of the mixture-of-distributions hypothesis (MDH) for returns that explicitly accommodates numerous heterogeneous information arrival processes.

An alternative to ARFIMA is the heterogeneous autoregressive (HAR) model suggested by Corsi (2009) (see also Aït-Sahalia and Mancini, 2008; Corsi et al., 2007). Extending the heterogeneous ARCH model of Müller et al. (1997), the long-memory pattern is reproduced by summing of (a small number of) volatility components constructed over different horizons. The basic ideas stems from the so called 'heterogeneous market hypothesis' presented by Müller et al. (1993), which recognized the presence of heterogeneity in traders. Differently from Andersen and Bollerslev (1997), in this latter view the multi-component structure in the volatility is to be found in the heterogeneity of agents rather than in the heterogeneous nature of the information arrival.

Consider the case with a single asset. Defining the k-period realized volatility component by the sum of the single-period realized volatilities, i.e.

$$\left(\sqrt{RV}\right)_{t-k:t-1} = \frac{1}{k}\sum_{j=1}^{k}\sqrt{RV_{t-j}}, \tag{2.2.8}$$

the HAR model for realized volatility of Corsi (2009), including the daily, weekly and monthly realized volatility components, is given by

$$\sqrt{RV}_t = \alpha_0 + \alpha_d + \sqrt{RV}_{t-1} + \alpha_w \left(\sqrt{RV}\right)_{t-5:t-1} + \alpha_m \left(\sqrt{RV}\right)_{t-22:t-1} + \mu_t. \tag{2.2.9}$$

In Corsi (2009) μ_t is assumed to be Gaussian white noise., whereas in Corsi et al. (2007), a standardized normal inverse Gaussian (NIG) is chosen to deal with the non-Gaussianity of the error terms.

In the spirit of the HAR model, we propose here to model the conditional realized covariance matrix Y_t with an autoregressive Wishart process which accounts for the temporal aggregation of the covariance matrix. We call this process HAR-WAR process. In the sequel, we will show that this process, can be interpreted as a particular WAR(23) process.

Define the k-period realized covariance matrix component by the sum of the single-period realized covariance matrices:

$$Y_{t-k:t-1} = \frac{1}{k}\sum_{j=1}^{K} Y_{t-j} \tag{2.2.10}$$

Combining a WAR(p) structure with the temporal aggregation induced by the HAR model, we write the process Y_t as:

$$Y_t = M_1 Y_{t-1} M_1' + M_2 Y_{t-5:t-1} M_2' + M_3 Y_{t-22:t-1} M_3' + K\Sigma + \eta_t, \tag{2.2.11}$$

Now, opening the summations and aggregating according to the same lag, we get:

$$\begin{aligned} Y_t \quad = \quad & \left(M_1 Y_{t-1} M_1'\right) + \left(\tilde{M}_2 Y_{t-1} \tilde{M}_2' + \tilde{M}_3 Y_{t-1} \tilde{M}_3'\right) + \cdots + & (2.2.12) \\ & \left(\tilde{M}_2 Y_{t-5} \tilde{M}_2' + \tilde{M}_3 Y_{t-5} \tilde{M}_3'\right) + \tilde{M}_3 Y_{t-6} \tilde{M}_3' + \cdots + \tilde{M}_3 Y_{t-22} \tilde{M}_3' + & (2.2.13) \\ & K\Sigma + \eta_t, & (2.2.14) \end{aligned}$$

with $\tilde{M}_2 = \frac{1}{\sqrt{5}} M_2$ and $\tilde{M}_3 = \frac{1}{\sqrt{22}} M_3$.

To interpret the process as a WAR(22), we simply rewrite it as:

$$Y_t = M_1 Y_{t-1} M_1' + \sum_{i=1}^{5} N_2 Y_{t-i} N_2' + \sum_{j=6}^{22} \tilde{M}_3 Y_{t-j} \tilde{M}_3' + K\Sigma + \eta_t. \tag{2.2.15}$$

where

$$N_2 \quad : \quad N_2 Y_t N_2' = \tilde{M}_2 Y_t \tilde{M}_2' + \tilde{M}_3 Y_t \tilde{M}_3'.$$

As for the WAR(p) process, the WAR-HAR process permits a *vech* representation, i.e.

$$vech(Y_t) = \sum_{j=1}^{22} A_j(M_1, M_2, M_3) vech(Y_{t-j}) + vech(K\Sigma) + vech(\eta_t) \qquad (2.2.16)$$

where $A_j(M_1 N M_2, \tilde{M}_3)$ is a matrix function of M_1, N_2 and $\tilde{M}_3$.

Since the HAR-WAR model is a WAR(22) characterized using only three autoregressive matrices, the reduction of the parametric space introduced in Section 2.2.3 is applied in this new context to matrices M_1, M_2 and M_3. This originates what we called the *full HAR-WAR*, the *diagonal HAR-WAR*, the *restricted diagonal HAR-WAR*, the *block HAR-WAR* and the *restricted block HAR-WAR*. The relations between block-structured models and causality restrictions presented in the previous section, are also valid for the HAR-WAR model.

2.3 Estimation

2.3.1 Identification

Following the exposition in Gourieroux et al. (2009), we obtain an analogous identification result for the block WAR and block WAR-HAR model. For ease of exposition we present only the estimation procedure for the WAR(1) process with diagonal autoregressive matrix M. The assumption of diagonal M, even if strict, renders the estimation extremely easy and fast. The extension to the diagonal HAR-WAR case is straightforward.

Under the assumption that $K > n - 1$ it is straightforward to show that:

i) K and Σ are identifiable while the autoregressive coefficients in M (an thus M_1, M_2 and M_3) are identifiable up to their sign.

ii) Σ is first-order identifiable up to a scale factor and M is first-order identifiable up to its sign. The degree of freedom K is not first-order identifiable but is second-order identifiable.

2.3.2 First-order identification

Following Gourieroux et al. (2009), the first-order conditional moments can be used to calibrate the parameters in M and Σ, up to the sign and scale factor, respectively.

As the first-order method of moments is equivalent to non-linear least squares, the estimator is defined as:

$$\left(\hat{M}, \hat{\Sigma}^*\right) = \text{Argmin}_{M,\Sigma^*} S^2\left(M, \Sigma^*\right)$$

where

$$\begin{aligned} S^2\left(M, \Sigma^*\right) &= \sum_{t=2}^{T}\sum_{i<j}\left(Y_{ij,t} - \sum_{k=1}^{n}\sum_{l=1}^{n} Y_{kl,t-1} m_{ik} m_{lk} - \sigma_{ij}^*\right)^2 \\ &= \sum_{t=2}^{T} \|vech(Y_t) - vech(MY_{t-1}M' + \Sigma^*)\|^2 \end{aligned}$$

and $\Sigma^* = K\Sigma$.

As mentioned in Gourieroux et al. (2009), any statistical software which accounts for heteroskedasticity can be used to obtain the estimates. We present here the complete procedure under the assumption that M is diagonal as we want to emphasize the quickness of the algorithm.

For each $Y_t, t = 1, \ldots, T$ of dimensions $n \times n$, we consider the matrix $\mathbf{Y}$, of dimensions $T \times n(n+1)/2$ build with the *vech* of Y_t for each time $t = 1, \ldots, T$; i.e. the i-th row of $\mathbf{Y}$ is $vech(Y_i)$.

Under the hypothesis that M is diagonal, define $a = diag(M)$ and $dg(a)$ as the diagonal matrix with the vector a as diagonal. Then

$$MY_{t-1}M' = dg(a)Y_{t-1}dg(a) = (aa') \odot Y_{t-1} \tag{2.3.1}$$

and

$$vech(MY_{t-1}M') = vech(aa') \odot vech(Y_{t-1}) \tag{2.3.2}$$

where $\odot$ denotes the elementwise product. Define $[\mathbf{Y}]_2^T$ as the matrix obtained from $\mathbf{Y}$ when dropping the last row, i.e. considering the time from T down to time 2. Define $A = vech(aa')$ and $Z = vech(\Sigma^*)$. The residual matrix W is obtained as

$$W = [\mathbf{Y}]_2^T - \left(A' \otimes i_{T-1}\right) \odot [\mathbf{Y}]_1^{T-1} - Z' \otimes \mathbf{i}_{T-1} \tag{2.3.3}$$

where $\mathbf{i}_{T-1}$ is a $T - 1 \times 1$ vector of ones and $\otimes$ denotes the Kronecher product.

Then the minimization problem reduces to:

$$\left(\hat{M}, \hat{\Sigma}^*\right) = \text{Argmin}_{M,\Sigma^*}\left[\mathbf{i}_{T-1}'\left(W \odot W\right)\mathbf{i}_{n(n+1)/2}\right]. \tag{2.3.4}$$

With our data set of four assets and 2,174 trading days (see Section 2.4 for a detailed description), only 1.2 seconds for the diagonal case (0.7 seconds for the restricted diagonal case) on a Pentium 4 PC are necessary to obtain the estimates. This result, if compared with the 42 seconds required from the same data set when a DCC model (Engle, 2002) is fitted, represents a great improvement.[3] For the diagonal HAR-WAR only 5 seconds are required, and for its restricted version only 3.9 seconds. See Table 2.8 for all the other specifications.

2.3.3 Second-order identification

Whereas the estimation of the entries of the autoregressive matrix M and of the innovation variance Σ (up to multiplication for a scale parameter) is relatively straightforward, the estimation of the degrees of freedom poses some challenges. We first present the estimation procedure introduced in Gourieroux et al. (2009) and then show how the same parameter K can be estimated relying on the fact that, given a portfolio allocation α, its volatility $\alpha' Y_t \alpha$ is gamma-distributed with a shape parameter equal to K.

Consider the simple WAR(1) model. The marginal distribution of the WAR(1) is the centered Wishart distribution, defined as $W(K, 0, \Sigma(\infty))$, where $\Sigma(\infty)$ is computed from

$$\Sigma(\infty) = M\Sigma(\infty)M' + \Sigma. \tag{2.3.5}$$

Thus, the conditional variance of a portfolio's volatility is given by:

$$V(\alpha' Y_t \alpha) = \frac{2}{K}[\alpha' \Sigma^*(\infty)\alpha]^2, \tag{2.3.6}$$

where α is a vector of dimension $(n \times 1)$ and $\Sigma^*(\infty) = K\Sigma(\infty)$. A consistent estimator of the degrees of freedom K can be computed as follows:

Step 1 Compute $\hat{\Sigma}^*(\infty)$ as solution of

$$\hat{\Sigma}^*(\infty) = \hat{M}\hat{\Sigma}^*(\infty)\hat{M}' + \hat{\Sigma}^*(\infty). \tag{2.3.7}$$

Step 2: Chose a portfolio allocation and compute its sample volatility

$$V(\alpha' Y_t \alpha) = \frac{1}{T}\sum_{t=1}^{T}\left[\alpha' Y_t \alpha - \frac{1}{T}\sum_{t=1}^{T}\alpha' Y_t \alpha\right]^2. \tag{2.3.8}$$

Step 3: A consistent estimator of K is:

$$K\hat{(\alpha)} = 2[\alpha' \hat{\Sigma}^*(\infty)\alpha]^2 / \hat{V}(\alpha' Y_t \alpha) \tag{2.3.9}$$

Step 4: A consistent estimator of Σ is $\hat{\Sigma}(\alpha) = \hat{\Sigma}^* / \hat{K}(\alpha)$.

A derivation of the above estimator for the general stationary WAR(p) process is reported in the Appendix.

[3] To ensure a fair benchmark, we tested both our Matlab code and the one provided by Kevin Sheppard in his UCSD toolbox.

This method provides consistent estimates of the degrees of freedom but is problematic in two aspects: first, it needs to estimate the matrix $\Sigma(\infty)$; second, it makes use of the estimates $\hat{M}$ and $\hat{\Sigma}$, carrying their estimation error into the estimate of $\hat{K}$.

A more direct way that does not need to rely on the estimates of M and Σ comes from the distribution of the volatility of a portfolio.

Consider a portfolio allocation $\alpha \in \mathbb{R}^n$. We know that the unconditional distribution of Y_t is a $W(K, 0, \Sigma(\infty))$, a centered Wishart distribution. We can therefore easily show[4] that

$$\alpha' Y_t \alpha \sim \text{Ga}\left(\frac{K}{2}, 2\alpha'\Sigma(\infty)\alpha\right), \tag{2.3.10}$$

i.e. the distribution of the portfolio with allocation α is a gamma distribution with the degrees of freedom K as shape parameter. An unbiased estimator of K can be obtained simply via maximum likelihood by fitting a gamma distribution to the process $\alpha' Y_t \alpha$[5]. As shown in Bonato (2009a), both estimators are unbiased but the second one is statistically more efficient. However, it is important to recall that these results are valid only if a WAR(1) is the true data generator process (DGP). This assumption, even if realistic, is far from being true, and a divergence in the values of the estimates is expected. In particular, Bonato (2009a) shows that in the presence of extreme observations or when the DGP is not a Wishart process, the estimates for the degrees of freedom using the WAR model are perceptibly lower than predicted by the theory via gamma distribution. A comparison of the two estimates should give a sort of measure of goodness of fit of the WAR model. A perfect fit should bring the two values to coincide.

The value of the degrees of freedom is the key element in determining whether the process is non-degenerate ($K \geq n$) or if it admits density ($K > n - 1$). Once the estimated degrees of freedom using the two estimators confirm the stationarity of the process, then the question of which estimator of K is to be used is no longer an issue, as the forecasted covariance matrices are independent of K. In fact, $\hat{M}$ and $\hat{\Sigma}^*$ are first-order identifiable and are only required to compute $E_t(Y_{t+1})$, as shown in Equation (3.2.4). Recall that $\hat{\Sigma} = \hat{\Sigma}^*/\hat{K}$ and K is second-order identifiable. So we do not need $\hat{K}$ to obtain $\hat{\Sigma}^*$.

2.4 The data

Our model introduces parametric restrictions by grouping the assets according to their type. For this reason we consider a portfolio composed of two currencies and two treasury bills. Bonds and currencies are in fact not likely to be correlated and thus our choice not to impose limited spillover between variances is justified *a priori*. As currencies we used USD/CHF and USD/GBP

[4]See, for example, the proof given in Meucci (2005, Technical Appendix, p. 33-34) or the Appendix of this paper.

[5]When performing the ML estimation one should be careful to the parametrization of the Gamma density function. According to Meucci's notation, it would be for instance $\alpha' Y_t \alpha \sim \text{Ga}(K, \alpha'\Sigma(\infty)\alpha)$

five-minute spot prices provided by Olsen and Associate Zürich . USD/CHF prices were available from 2 January 1997 to 9 August 2005 and USD/GBP series was covering the period from 2 January 1997 to 31 October 2006. The second group consists of the prices of the 10-year and 30-year U.S. treasury bills. These futures are traded at the Chicago Board of Trade (CBoT) from 7:20 to 14:00 Eastern Standard Time (EST). Our samples contain five-minute prices from 2 January 1997 to 29 June 2007. We adopted the conventional[6] practice of using the futures contract with the largest trading volume. As the contract approached maturity (five trading days before), we moved to the next contract, ensuring no overlapping periods in the price sequence and no returns computed on prices from different contracts. Days in which at least one of the series had no match with the other three (e.g. when the CBoT was closed) were dropped. In addition, 23 October 1997, 9 September 1998, 14 April 2003 and 11 October 2004 were removed from the sample due to the presence of irregularities. This left us with 2,147 trading days.

Table 2.1: Summary statistics of five-minute and daily returns. Daily returns are computed as the logarithm of the difference between the closing price and opening price multiplied by 100. Exchange rates are traded round the clock but as we are interested in a portfolio, only the trading hours coinciding with the CBoT trading hours were considered.

Return		CHF/USD	GBP/USD	T-10Y	T-30Y
5-min	Mean	0.0003	-0.0004	0.0001	0.0001
	Maximum	1.2716	0.6765	0.7856	0.7916
	Minimum	-1.3690	-0.6763	-1.0124	-0.8992
	St. dev.	0.0575	0.0433	0.0570	0.0367
	Skewness	-0.0322	-0.0145	-0.3391	-0.4123
	Kurtosis	16.1390	10.9153	11.1789	19.1486
Daily	Mean	-0.0250	-0.0277	0.0049	0.0076
	Maximum	3.1195	1.4240	1.9022	1.0802
	Minimum	-2.8374	-2.0079	-1.9112	-1.3626
	St. dev.	0.4967	0.3403	0.4970	0.3199
	Skewness	-0.1294	-0.0722	-0.3460	-0.3030
	Kurtosis	5.3625	4.8464	3.9230	4.2370

Currencies are traded around the clock. T-bills are traded during the CBoT trading day and virtually round the clock on GLOBEX starting from 1 July 2003. As our samples start in 1997 we studied only the overlapping trading hours, i.e. the trading hours of the CBoT. To remove the overnight effect we did not consider the first 15 minutes after the opening. Table 2.1 reports the descriptive statistics for the five-minute and daily returns for the four asset we considered. Intraday returns were constructed taking the first differences of the log-prices and multiplying by 100. Daily returns were computed as the logarithm of the difference between the closing price and opening price multiplied by 100. The typical stylized facts are observed: negative skewness,

[6]As done in Martens and van Dijk (2007) and de Pooter et al. (2006) among others.

excess of kurtosis in both daily and intraday T-bills returns and skewness close to zero for the exchange rates.

The trading day we constructed runs from 7:40 (first observation) to 14:00 (last observation), resulting in 76 five-minute returns which we used to construct the series realized covariance matrices. Descriptive statistics for the realized volatilities of the four assets are reported in Table 2.2. Figure 2.1 shows the realized volatilities estimated from the data. The evolution of the realized correlation is presented in Figure 2.2

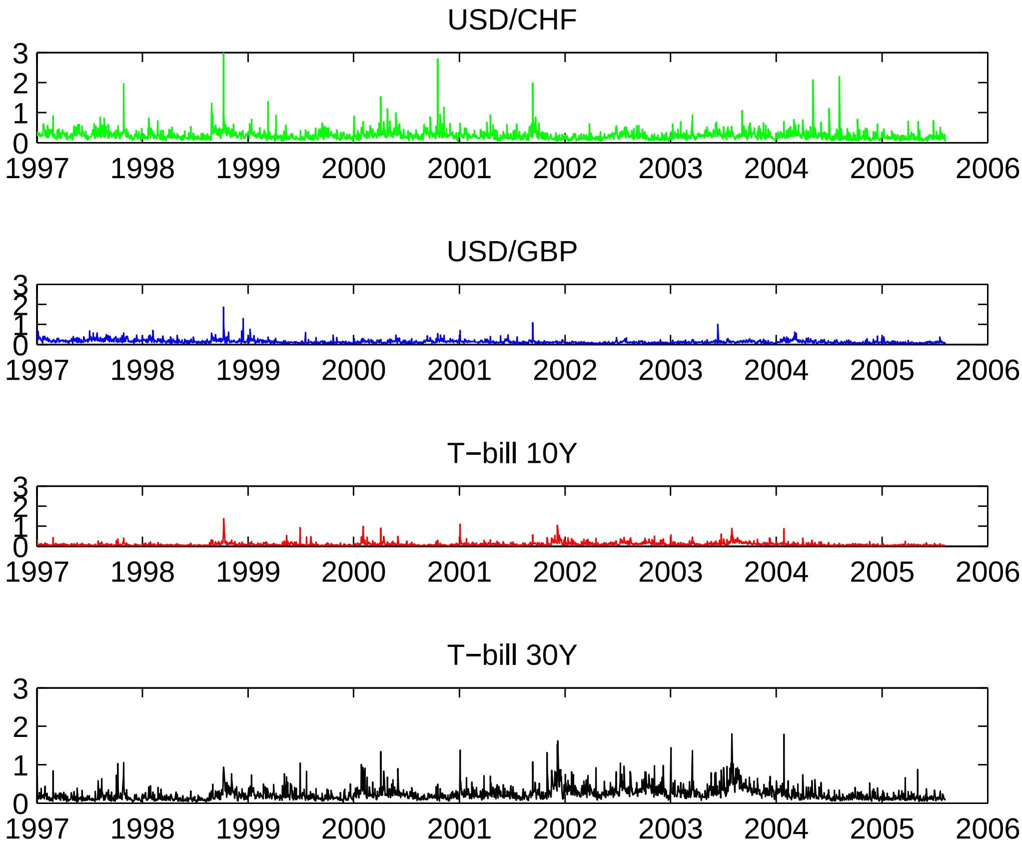


Figure 2.1: Daily realized volatilities for the two exchange rates and the two treasury bonds.

In the next step we constructed the series of realized covariance matrices using the classical estimator presented in Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2004) and used, for example, in de Pooter et al. (2006):

$$Y_t = \sum_{i=1}^{I} r_{t-1+ih,h} r'_{t-1+ih,h} \tag{2.4.1}$$

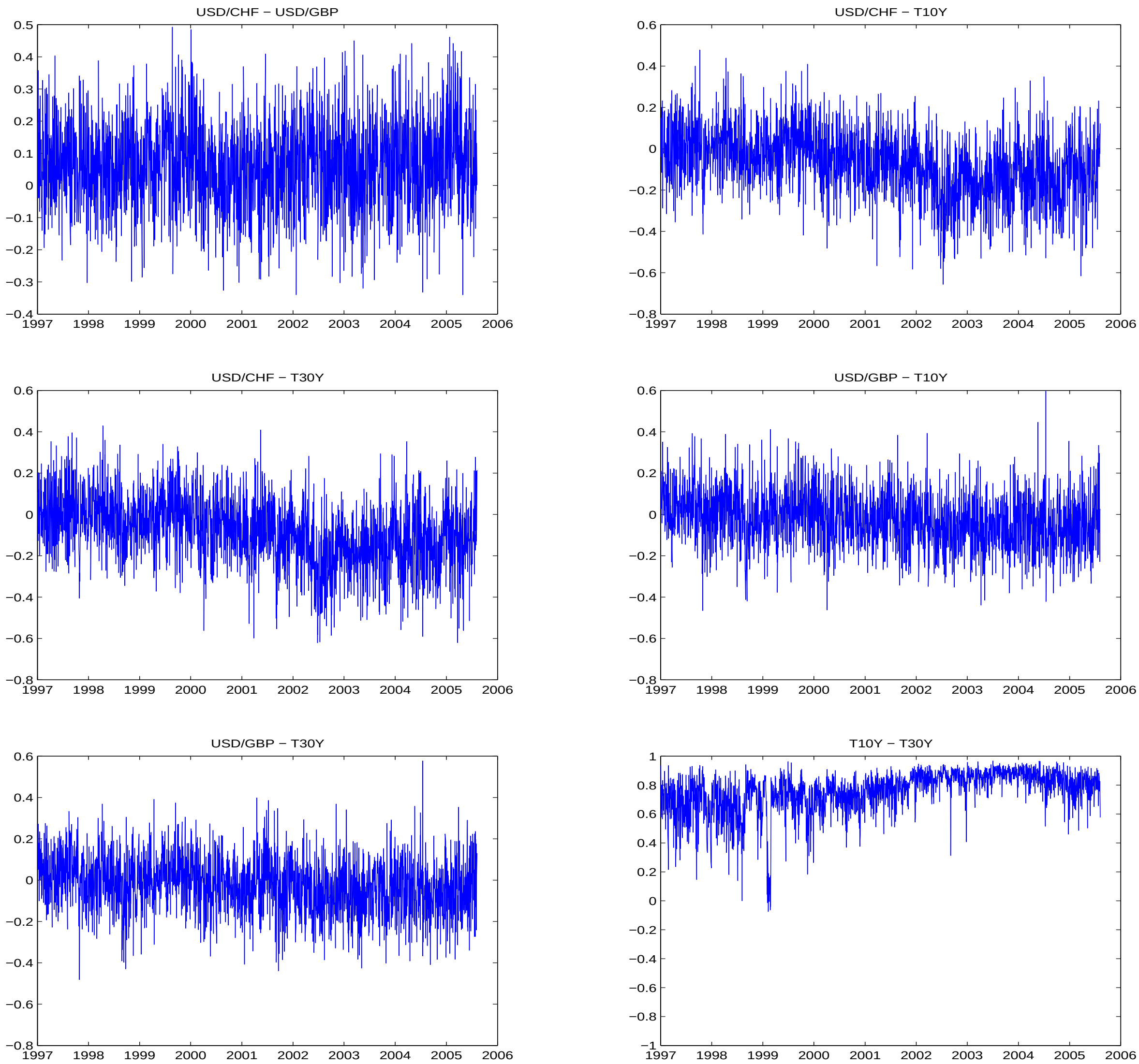
We indicate with Y_t the realized covariance matrix at time t in order to to be coherent with our previous notation and because the use of Σ would probably create confusion as Σ denotes the covariance matrix of the Gaussian vector underlying the WAR(1) model. $r_{t-1+ih,h} \equiv p_{t-1+ih} - p_{t-1+(i-1)/h}$ denotes the $(n \times 1)$ vector of returns for the i-th intraday period on day t, for $i = 1, \ldots, I$, and with $n = 4$ the number of assets. I is the number of intraday intervals, each of length $h \equiv 1/I$. In our case, with a frequency of five minutes, $I = 76$. One shortcoming of the covariance matrix estimator we adopted is that it is not efficient in the presence of market microstructure noise and asynchronous trading (see for example Sheppard, 2006, Lunde and Voev, 2007, Barndorff-Nielsen et al., 2008b, Mancino and Sanfelici, 2008, among others). We think this does not represent an issue here as, first, we use very liquid assets that are traded in the same markets (CBoT for the futures and OTC for the currencies). This reduces the distortion induced by stale prices, non-homogenous trading time, data points irregularly spaced, asynchronism, different institutional features using different trading platforms or exchange systems. Secondly, as shown in Barndorff-Nielsen et al. (2008b) using intraday data of 10 stocks from the Dow Jones index, the estimated realized covariance matrices based on 5-minute returns are not significantly biased[7] (compared to the matrices constructed using the outer products of the open to close returns) even though realized kernels remain the preferred estimators. In contrast to de Pooter et al. (2006) we did not consider overnight returns. Including overnight returns would affect only the volatility of the T-bills because currencies are traded 24 hours and their equivalent to the overnight returns would be the over-weekend return. Therefore we contend that adding overnight returns to only some components of the portfolio would induce distortion in the realized volatility of the portfolio itself.

Table 2.2: Summary statistics for the realized volatilities

Realized volatility	CHF/USD	GBP/USD	T-10Y	T-30Y
Mean	0.2511	0.1422	0.2466	0.1022
Maximum	2.9772	1.8661	1.8043	1.3761
Minimum	0.0184	0.0164	0.0276	0.0119
St. dev.	0.1856	0.1039	0.1895	0.1006
Skewness	5.5066	4.8388	2.6636	4.5772
Kurtosis	59.7536	54.3341	14.2783	37.2670

[7] For a given estimator, say $Y_t = Cov_t^{5m}$, Barndorff-Nielsen et al. (2008b) consider the difference $d_t = Cov_t^{5m} - Cov_t^{OtoC}$ where Cov_t^{OtoC} is the outer products of the open to close returns, which when averaged over many days provide an estimator of the average covariance between asset returns. The sample bias is computed as $\bar{d}$ and the robust variance as $\bar{e}^2 = \gamma_0 + 2\sum_{h=1}^{q}\left(1 - \frac{h}{q+1}\right)\gamma_h$, where $\gamma_h = \frac{1}{T-h}\sum_{t=1}^{n-h}\eta_t\eta_{t-h}$. Here $\eta_t = d_t - \bar{d}$ and $q = \text{int}\{4(T/100)^{2/9}\}$. Under the null hypothesis of no difference between the two estimators at one percent level $\left|\sqrt{T}\bar{d}/\bar{e}\right| < 2.326$ for each entry of Cov_t^{5m}.

Figure 2.2: Evolution of the realized correlations for the four assets in analysis.



2.5 Empirical application

2.5.1 Estimation results

The first model we estimated is the full WAR(1), in which the matrix M is full. The estimates of the entries of M and Σ are reported in Table 2.3 and 2.4, respectively. As shown in Equation (2.2.7), the impact of the past values of realized variances and covariances on future realized variances and covariances is a function of the entries of M, so, rather than checking the significance of the elements of M, we are interested in checking the significance of the coefficients $a_i, b_i, c_i,\ i = 1, \ldots, 3$, i.e. the coefficients that directly affect the realized variance-covariance matrix forecasts.

0.4044	0.1033	0.0764	-0.1442
(3.3985)	(0.2273)	(0.1868)	(-0.2282)
-0.0602	0.5637	-0.0344	0.0600
(-0.2441)	(4.2327)	(-0.1067)	(0.1235)
0.0323	0.0008	0.7204	-0.1047
(0.2425)	(0.0003)	(3.3614)	(-0.3092)
-0.0128	0.0489	0.1753	0.4037
(-0.0715)	(0.2063)	(0.5773)	(0.9577)

Table 2.3: Estimated latent autoregressive matrix M for the full WAR(1) model. t-ratios in parenthesis.

0.0424	0.0007	-0.0011	0.0002
(7.9627)	(0.1110)	(-0.1812)	(0.0445)
	0.0197	-0.0017	-0.0023
	(3.7092)	(-0.3023)	(-0.4465)
		0.0279	0.0136
		(4.8620)	(2.7554)
			0.0123
			(2.8124)

Table 2.4: Estimated latent autoregressive matrix Σ for the full WAR(1) model. t-ratios in parenthesis.

Table 2.5 reports the estimates and the t-test values of the parameters that determine the best prediction of Y_t as given by a WAR(1) model. For simplicity we will only consider the case of two assets and report the estimates of the different pairs of combinations of the two currencies and two T-bills we used in our analysis. The parameter a_1, which tells us the effect of the realized volatility at time $t-1$ on the realized volatility expected at time t, is significant for all the pairs[8] . We have the same results for the coefficients b_2 and c_3, the autoregressive parameters for the realized covariances and realized variances of the second component of the pair. The only exceptions are the couples CHF-GBP and T30-T10. In particular, for the latter pair, only the autoregressive coefficient for the 30-year U.S. treasury bill is statistically significant.

It is very important to note that the rest of the coefficients are not statistically significant for any of the different combinations of pairs. This suggests that a reduction of the parameters of the models hypothesizing a limited spillover is reasonable and to some extent necessary.

The estimates of the autoregressive matrix M and the covariance matrix Σ for the four specifications of the WAR(1) model, the diagonal, the diagonal restricted, the block-diagonal and the restricted block-diagonal are reported in Tables 2.6 and 2.7. Standard errors are in parenthesis. Starting at the top left of Table 2.6, we see that imposing the same value of the autoregressive coefficient for assets belonging to the same type is a sensible choice. Consider the diagonal WAR case. For the first two elements of the diagonal (exchange rates), we have a

[8]Recall from (2.2.7) that $a_1 = m_{11}^2$ so that the significance test is a one-sided test with 10% level at 1.28.

	CHF-GBP	CHF-T30	CHF-T10	GBP-T30	GBP-T10	T30-T10
a_1	**0.1613**	**0.1786**	**0.1806**	**0.3279**	**0.3364**	**0.5419**
	(1.5543)	(2.1789)	(2.1754)	(2.2469)	(2.2960)	(1.7310)
a_2	-0.0418	0.0130	-0.0027	0.0081	0.0196	0.1304
	(-0.4640)	(0.2369)	(-0.0340)	(0.0857)	(0.1500)	(0.5772)
a_3	0.0108	0.0009	0.0000	0.0002	0.0011	0.0314
	(0.2190)	(0.1184)	(0.0170)	(0.0429)	(0.0750)	(0.2874)
b_1	-0.0835	0.0260	-0.0054	0.0162	0.0392	0.2607
	(-0.3802)	(0.2363)	(-0.0338)	(0.0857)	(0.1500)	(0.5638)
b_2	0.2051	**0.2783**	**0.2629**	**0.3783**	**0.3627**	0.2722
	(1.2183)	(4.0238)	(3.2827)	(4.1564)	(3.4078)	(0.7471)
b_3	-0.1171	0.0406	-0.0078	0.0187	0.0421	0.1417
	(-0.4521)	(0.2365)	(-0.0340)	(0.0856)	(0.1491)	(1.3579)
c_1	0.0412	0.0004	0.0040	0.0001	0.0013	0.0161
	(0.2635)	(0.0799)	(0.1065)	(0.0417)	(0.0830)	(0.1876)
c_2	0.1143	-0.0137	-0.0389	0.0062	0.0224	-0.0507
	(0.5561)	(-0.1598)	(-0.2135)	(0.0833)	(0.1662)	(-0.3345)
c_3	**0.3173**	**0.4356**	**0.3815**	**0.4361**	**0.3883**	0.1602
	(1.9316)	(5.4035)	(2.4987)	(5.2972)	(2.5243)	(0.4589)

Table 2.5: Estimates and t-ratios for the coefficients of Equation (2.2.7). Coefficients that are significant at the 10% level are shown in bold.

common parameter 0.4585 against 0.4175 and 0.5636. For the T-bills we have an autoregressive parameter for the volatilities equal to 0.6481 in front of 0.6583 and 0.6209. Including spillover between assets belonging to the same sector affects only the autoregressive parameter of the 30-years T-bill and appears unnecessary as most of the off-diagonal coefficients are not significant at the 5% level, confirming the findings reported in Table 2.5. The restricted block diagonal case presents estimates that are not compatible with the previous cases and this seems to suggest that this kind of specification might be too restrictive to model the covariance matrix. The estimation results for the HAR-WAR process are similar to those for the WAR process and are available upon request.

The estimated values for the degrees of freedom are reported in Table 2.8. To obtain the estimates the following allocation was used: $\alpha = (1\ 1\ 1\ 1)'$. Different allocations led to analogous results.

All the different specifications result in a number of degrees of freedom strictly bigger than n, $n = 4$ being the number of assets making up the portfolio, and thus the Wishart process is stationary and non-degenerate. All the estimates of K are close to each other except for the restricted block WAR-HAR. The resulting degrees of freedom equal to 6.5 are slightly bigger than in the other cases and this might be due to some problem in the optimization routine. Further investigation in this direction is necessary.

In addition to the estimated degrees of freedom, Table 2.8 also reports the number of pa-

Block WAR				Restricted block WAR			
0.4080	0.1060			0.2740	0.2740		
(3.5332)	(0.2383)			(4.8680)			
-0.0648	0.5626			0.2740	0.2740		
(-0.2649)	(4.2528)						
		0.7216	-0.1078			0.3282	0.3282
		(3.3565)	(-0.3175)			(12.8269)	
		0.1716	0.4035			0.3282	0.3282
		(0.5640)	(0.9389)				
Diagonal WAR				**Restricted diagonal WAR**			
0.4175				0.4584			
(4.2792)				(5.9889)			
	0.5636				0.4584		
	(4.4107)						
		0.6583				0.6481	
		(11.1432)				(13.595)	
			0.6209				0.6481
			(6.0008)				

Table 2.6: Estimated latent autoregressive matrix M for the different specification of the WAR(1) model. t-ratios in parenthesis.

Block WAR				Restricted block WAR			
0.0424	0.0007	-0.0002	-0.0003	0.0451	-0.0049	0.0000	0.0000
(8.0529)	(0.1140)	(-0.0604)	(-0.0828)	(11.1560)	(-1.2006)	(0.0070)	(0.0069)
	0.0197	-0.0019	-0.0014		0.0228	-0.0024	-0.0016
	(3.7136)	(-0.6149)	(-0.4363)		(5.6421)	(-0.7805)	(-0.4998)
		0.0279	0.0136			0.0371	0.0127
		(4.8738)	(2.7514)			(9.9012)	(3.3877)
			0.0124				0.0076
			(2.8801)				(2.0225)
Diagonal WAR				**Restricted diagonal WAR**			
0.0424	0.0011	-0.0004	-0.0004	0.0406	0.0011	-0.0004	-0.0004
(8.1190)	(0.3423)	(-0.1238)	(-0.1264)	(8.5055)	(0.3536)	(-0.1201)	(-0.1201)
	0.0198	-0.0019	-0.0014		0.0230	-0.0021	-0.0015
	(3.7920)	(-0.6012)	(-0.4396)		(6.1516)	(-0.6732)	(-0.4760)
		0.0285	0.0154			0.0292	0.0151
		(5.6888)	(4.2106)			(6.6184)	(4.2865)
			0.0128				0.0121
			(3.1117)				(3.5788)

Table 2.7: Estimated latent autoregressive matrix Σ for the different specification of the WAR(1) model. t-ratios in parenthesis.

Specification	Parameters	CPU time (secs)	$\hat{K}$	fval	Ranking
full WAR	27	117	4.8	209.01	9
block diag. WAR	19	94	4.9	209.11	8
restr. block diag. WAR	13	21	4.8	231.96	5
diagonal WAR	15	1.22	4.8	209.39	2
restr. diag. WAR	13	0.71	4.8	209.80	1
full HAR-WAR	59	531	4.7	189.78	11
block diag. HAR-WAR	35	410	4.7	189.37	10
restr. block diag. HAR-WAR	17	92	6.5	198.52	7
diagonal HAR-WAR	23	3.5	4.6	187.45	4
restr. diag. HAR-WAR	17	2.5	4.7	187.54	3
DCC	14	42	-	-	6
diag. BEKK	18	639	-	-	12
$\hat{K}$ via gamma dist.			7.09	s.e. (0.8)	

Table 2.8: Estimate of the degrees of freedom for the different specifications of the WAR and HAR-WAR models (last column). The first column reports the number of parameters for each specification. The CPU necessary to obtain the estimates are reported in the second column. *fval* is the value of the function (2.3.4) at the minimuim.The last row reports the value of K when it is estimated relying on the gamma distribution for the variance of the portfolio.

rameters for each model and the CPU time necessary to obtain the estimates on a Pentium IV PC. The advantage of using a diagonal model (either WAR or HAR), compared with the full counterpart, is notable. The time required to obtain the estimates ranges from 0.71 to 5 seconds, a great improvement compared, for example, with the 323 seconds required by the diagonal BEKK of Engle and Kroner (1995), which assumes the same autoregressive structure for the latent variance-covariance matrix[9].

2.5.2 Variance Forecasting

The ability to forecast the volatility of a financial position is a key factor in many activities like risk management, portfolio optimization or option pricing, just to mention the most common. For this reason we preferred to give more emphasis to the out-of-sample forecast of the proposed model, rather than the in-sample fit and in-sample forecast. Of course, in-sample fit is important to determine the goodness of a model; however, unreported results showed that the WAR models have a very poor in-sample forecasting ability. Our suspicion is that the degrees of freedom are unlikely to be constant through time, and therefore fitting the model to the entire series is not appropriate. To check the variation of the degrees of freedom within the sample, we adopt a rolling window of 21 trading days to recursively estimate the WAR model. Figure 2.3 shows the values of the estimated degrees of freedom computed using the classical estimator as in Gourieroux (2007) (red line) and the estimator that relies on the gamma distribution (blue line). We can clearly

[9] Again, to estimate the parameters of the BEKK model we used the Matlab code provided by Kevin Sheppard in the UCSD Garch toolbox.

see that the degrees of freedom are far to be constant over time and that the values obtained relying on the gamma distribution are generally higher than the ones obtained using the classical estimator. Plotted is also the volatility of a portfolio (green line) we built with the 4 assets for our forecasting exercise. There seems to be a relation between the degrees of freedom and the realized volatility of the portfolio. In fact, high peaks in the volatility series coincides with lower values for $\hat{K}$, especially when the classical estimator is used. This is in line with the findings of Bonato (2009a) where it is shown that extreme observations in the variance-covariance process result in lower estimated degrees of freedom.

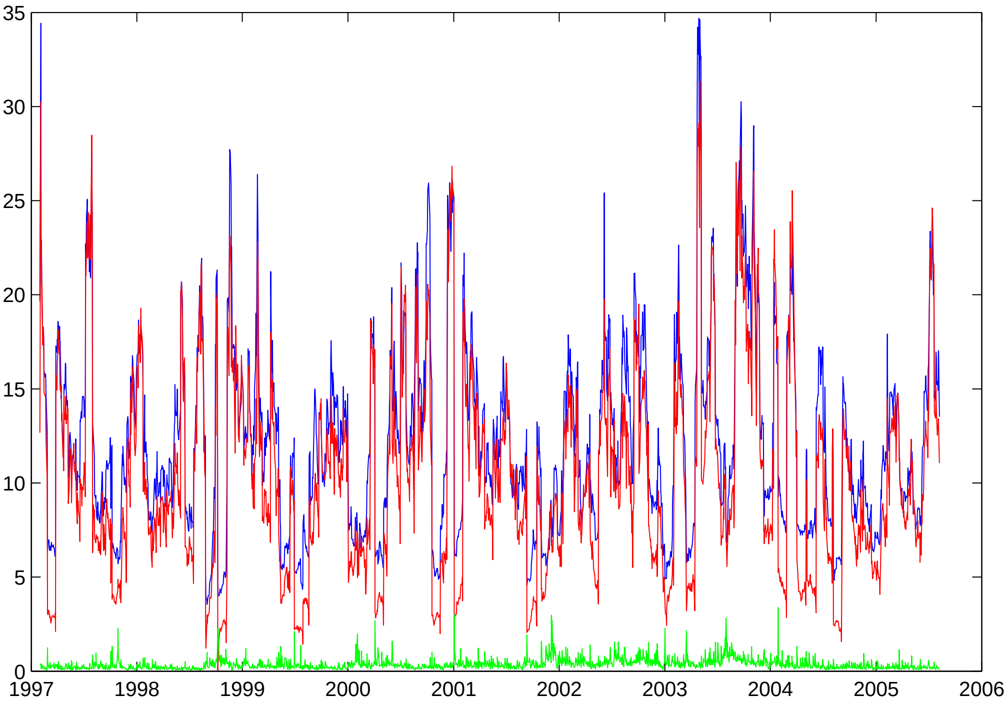


Figure 2.3: Estimated degrees of freedom using the classical estimator of Gourieroux (2007) (red line) and the estimator that uses the gamma distribution (blue line) when a rolling window of 21 trading days is used. The green line represents the realized volatility of the portfolio built with the 4 assets from day 21 until the end of the sample.

Our first step in this forecasting exercise is to construct a portfolio with the series of two exchange rates and two treasury bills. We assume that the value of the portfolio is in dollars and that it therefore carries a long position for the treasury bills and a short position in currencies. For simplicity, we assume equal (positive) weights for the treasury bills and equal (negative) weights for the exchange rates. In particular, we assume that the owner of the portfolio invests 0.75 of his wealth for each of the T-bills and short-sells 0.25 for each of the currencies to buy CHF and GBP against USD, respectively. The forecasting period runs from 2 January 2003 until 8 August 2005, resulting in 653 one-step-ahead forecasts. For each day the realized variance of the portfolio is forecast by fitting a WAR model to the series of covariance matrices and re-estimating the model at each step. As already mentioned above, the degrees of freedom are likely not to be constant and therefore at each step the model was estimated using a rolling window of 100 trading days, as done in Aït-Sahalia and Mancini (2008). Table 2.9 presents the results of the Mincer-Zarnowitz regression:

$$IV_t^{1/2} = b_0 + b_1 E_{t-1}[RV_t^{1/2}] + \text{error}, \tag{2.5.1}$$

where IV_t is the realized volatility of the portfolio at time t and $E_{t-1}[RV_t]$ is the forecasted realized volatility. Standard errors are reported in parenthesis. The R^2 across the models varies from 0.3209 for the full WAR(1) to the 0.3655 for the diagonal HAR-WAR. The moving windows estimation of the various WAR models delivered acceptable R^2, that are, for instance, slightly higher than those reported in Andersen et al. (2003).

It is interesting to note that the full WAR(1) model has a worse performance if compared with its restricted counterparts. This might be due to the fact that the full model is not the most appropriate as it carries over the estimation error of the parameters into the forecasts, which means that it is not as good as a more parsimonious model. It should also be noted that, in terms of R^2, the difference between the diagonal model and the restricted diagonal model is not relevant. Neither is the difference between the block diagonal and the restricted block diagonal. The diagonal model has the highest R^2. This suggests that this simple parametrization is sufficient to capture the dynamics of the variances and covariances.

2.5.3 Distribution of the portfolio's realized volatility

As demonstrated in the Appendix, under the WAR hypothesis the realized volatility of a portfolio follows a gamma distribution with shape parameter $K/2$, where K denotes the degrees of freedom of the Wishart process and scale parameter $2\omega'\Sigma(\infty)\omega$ with $\Sigma(\infty)$ solution of

$$\Sigma(\infty)^* = M\Sigma(\infty)^* M' + \Sigma^*.$$

as in (2.3.7), where ω is the vector of portfolio weights, i.e. $\omega = [-.25\ -.25\ .75\ .75]'$. Figure 2.5 (left) displays the density of the realized volatility of the portfolio under the hypothesis that it follows a gamma distribution. The dashed red line represents the kernel density of the portfolio's realized volatility. The green dash-dot line is the density of a $Ga(K_\Gamma/2, 2\omega'\Sigma(\infty)\omega)$ where K_Γ denotes the degrees of freedom estimated via the gamma distribution. The blue line is the density of a gamma distribution but with K estimated as in Gourieroux et al. (2009), Steps 1-4. Recall that to obtain both the estimates for K $\alpha = (1\ 1\ 1\ 1)'$ was used.

In Figure 2.5 (right) we fitted a gamma distribution to the realized volatility of our portfolio. The blue line represents the kernel density of the realized variance, the blue line is the gamma fitting and the black dash dot line represents the log-normal density. Numerous studies (Andersen

	b_0	b_1	R^2
full WAR(1)	0.0226	0.8988	0.3209
	(0.0333)	(0.0512)	
block diagonal WAR(1)	0.0004	0.9349	0.3262
	(0.0342)	(0.0526)	
restr. block diag. WAR(1)	0.0046	0.9405	0.3224
	(0.0341)	(0.0524)	
diagonal WAR(1)	0.0064	0.9434	0.3299
	(0.0343)	(0.0526)	
restr. diag. WAR(1)	0.0059	0.9428	0.3298
	(0.0342)	(0.0526)	
full HAR-WAR	0.1387	0.7361	0.3103
	(0.0275)	(0.0429)	
block diag. HAR-WAR	0.0685	0.8439	0.3584
	(0.0284)	(0.0442)	
restr. block diag. HAR-WAR	0.0647	0.8440	0.3623
	(0.0284)	(0.0438)	
diagonal HAR-WAR	0.0520	0.8630	0.3662
	(0.0289)	(0.0446)	
restr. diag. HAR-WAR	0.0550	0.8594	0.3655
	(0.0286)	(0.0443)	

Table 2.9: Out-of-sample one-day-ahead forecast of $IV^{1/2}$. The models are estimated on a rolling window of 100 days from 2 January 2003 to 8 August 2005. Standard errors in parenthesis.

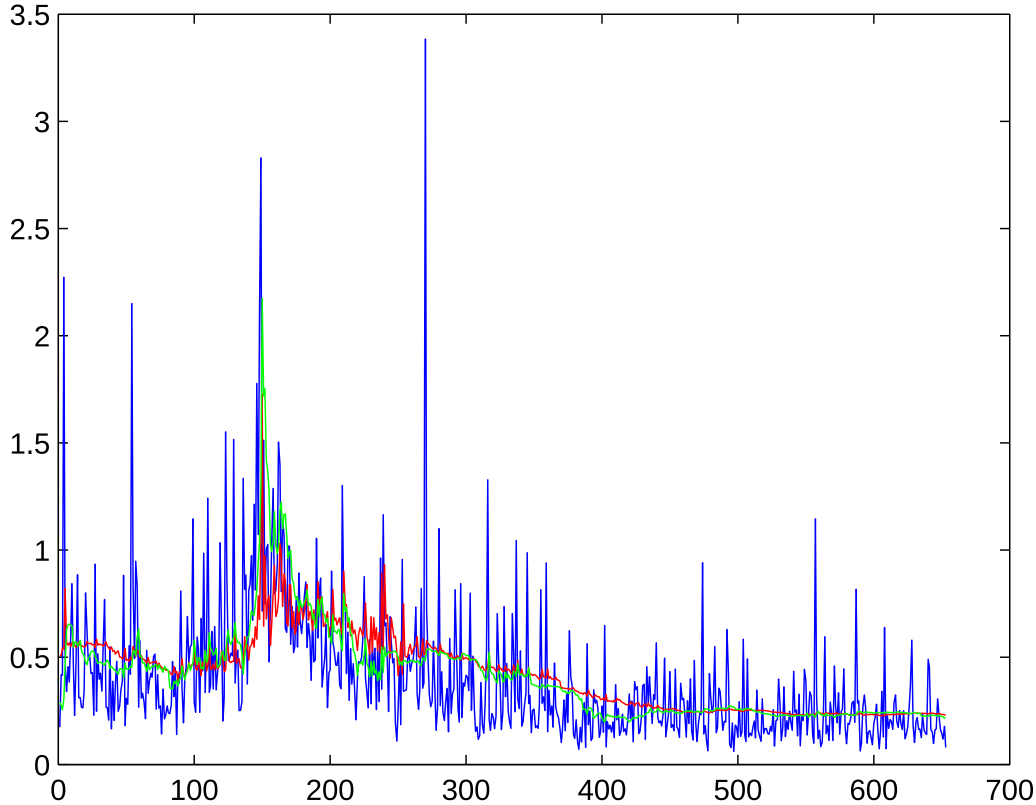


Figure 2.4: Out-of-sample forecast of the realized variance for the restricted diagonal WAR(1) (red line) and the restricted diagonal HAR-WAR model (green line). The blue line represents the ex-post observed realized volatility of the portfolio.

et al., 2003, among others) show that the logarithm of the realized volatility tends to follow a normal distribution. Is therefore no surprising that a lognormal distribution clearly better fits the distribution of the realized volatility of the portfolio when compared to a gamma distribution. On the other hand, the fit provided by the Wishart model, i.e. a gamma distribution, from a very rough graphical analysis, provides an acceptable alternative[10].

2.5.4 Value-at-Risk performance evaluation

Given the growing need to manage financial risk, risk prediction plays an increasing role in banking and finance. The Value-at-Risk (VaR) concept has emerged as the most prominent measure of downside market risk. Regardless of the criticisms levelled at it, regulatory requirements are heavily geared towards VaR. In the light of the practical relevance of the VaR concept, the need

[10]The assumption of a gamma distribution to model the realized volatility is also at the basis of the multiplicative model of Engle and Gallo (2006)

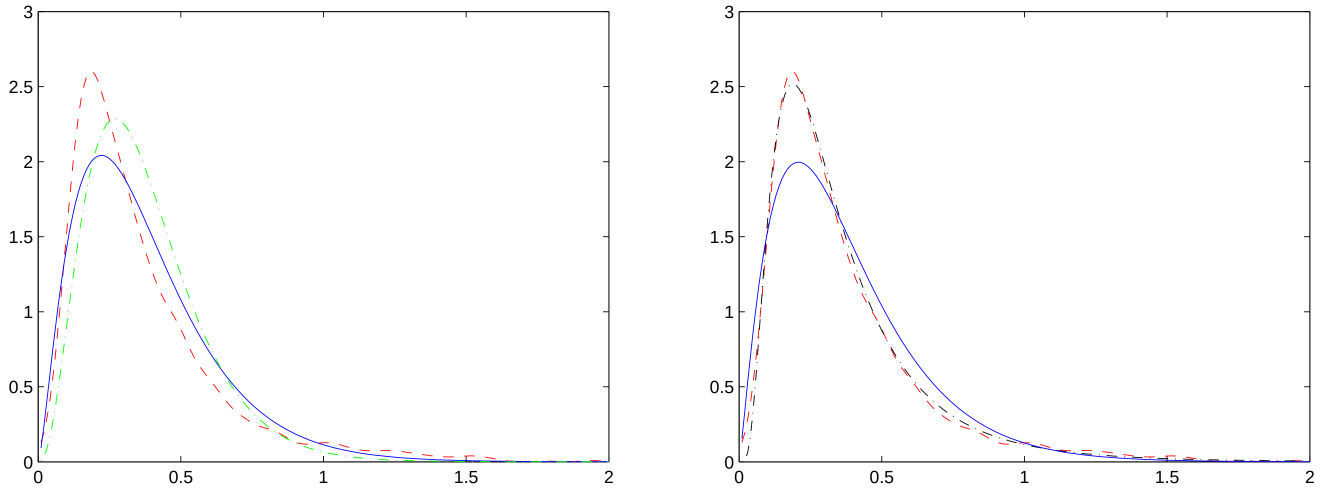


Figure 2.5: Kernel densities of the realized volatility of the portfolio (red dashed line), density of a $Ga(K_\Gamma/2, 2\omega'\Sigma(\infty)\omega)$ where K_Γ denotes the degrees of freedom estimated via the gamma distribution (blue line) and density of a gamma distribution (green dash-dot line) with K estimated as in Gourieroux et al. (2009) [left-hand panel]. Kernel densities of the realized volatility of the portfolio (red dashed line), gamma (blue line) and log-normal (black dash-dot) distribution fitted to the series [right-hand panel].

for reliable VaR estimation and prediction strategies arises. A key ingredient when predicting the VaR of a financial position is the ability to forecast the conditional variance of the asset considered. To fully test the proposed model we also consider VaR as an economic criterion to judge the forecast performances. We follow the methodology proposed in Giot and Laurent (2004), that, to our knowledge is the only paper, along with that by Andersen et al. (2003), Clements et al. (2008) and Brownlees and Gallo (2008), to deal with VaR and realized volatility.

A series of asset returns $r_t,\ t = 1, \ldots, T$, known to be conditionally heteroskedastic, is modeled as follows:

$$r_t = \mu_t + \epsilon_t \tag{2.5.2}$$

$$\epsilon_t = \sigma_t \nu_t \tag{2.5.3}$$

$$\mu_t = c(\eta|\Omega_{t-1}) \tag{2.5.4}$$

$$\sigma_t = h(\eta|\Omega_{t-1}), \tag{2.5.5}$$

where $c(\cdot, \Omega_{t-1})$ and $h(\cdot, \Omega_{t-1})$ are functions of Ω_{t-1} (the information set at time $t-1$), and depend on an unknown vector of parameters η; ν_t is an independent and identically distributed (i.i.d.) process, independent of Ω_{t-1}, with $E[\nu_t] = 0$ and $E[\nu_t^2] = 1$. μ_t is the conditional mean of r_t and σ_t is its conditional variance. In our setting we assume, for simplicity, a constant mean for all the assets in our portfolio. In particular, if r_t represents the return of the portfolio, $\mu_t = \mu$

and for the (realized) variance of the portfolio we have:

$$RV_t = \omega' Y_t \omega, \tag{2.5.6}$$

where ω are the portfolio weights as previously chosen. To compute one-day-ahead forecasts for the VaR of the daily return r_t using the conditional realized volatility, we re-estimate the model in Eq. (2.5.2) with constant conditional mean while the conditional variance is proportional to $RV_{t|t-1}$, the one-step-ahead forecast of the realized volatility of the portfolio; i.e. $\sigma_t^2 = \sigma^2 RV_{t|t-1}$(with σ^2 being an additional parameter to be estimated). σ^2 is used to ensure that the rescaled innovations have unit variance.

We used the same forecasting period as in the previous section. For each model we computed the one-day-ahead variance and then the one-day-ahead forecast of the VaR. A Gaussian distribution and a Student's t distribution were used to model the residuals z_t. Table 2.10 presents the performances of the different models in terms of VaR predictions. Forecasts of VaR at level $\rho = 1\%, 5\%$ and 10% were computed. For each model and distribution for ν_t, we reported the percentage of violations, i.e. the percentage of times that the realized return is smaller that the forecasted VaR. A good density forecast should satisfy two criteria. First, for a given VaR level ρ, the percentage of violations should be ρ. Second, violations should conditionally unpredictable, i.e. a violation of nominal ρ_1 VaR today should convey no information as to whether nominal ρ_2 percent VaR will be violated tomorrow.

To check the robustness of the different WAR models in this VaR forecast evaluation, we also report in Table 2.10 the p-values of the test proposed in Berkowitz (2001) to evaluate a density forecast. This test relies on the fact that for a given daily return r_t, if the series of one-day-ahead conditional density forecasts $\hat{f}_{t|t-1}(r_t)$ coincides with $f(r_t, \mathbb{I}_{t-1})$, it then follows under weak conditions that the sequence of probability integral transformation of r_t with respect to $\hat{f}_{t|t-1}(\cdot)$

$$u_t = \int_{-\infty}^{r_t} \hat{f}_{t|t-1}(s)ds = \widehat{F}(r_t) \tag{2.5.7}$$

should be i.i.d. uniformly distributed on (0,1). This transformation was first presented in Rosenblatt (1952).

If the series of u_t is distributed as an i.i.d. U(0,1), then

$$z_t = \Phi^{-1}\left[\int_{-\infty}^{r_t} \hat{f}_{t|t-1}(s)ds\right] \quad \text{is an i.i.d. N(0,1).}$$

Once the series has been transformed, it is straightforward to calculate the Gaussian likelihood and construct the likelihood ratio (LR) statistics.

In particular, Berkowitz (2001) suggested a test that allows the user to intentionally ignore model failures that are limited to the interior of the distribution; the proposed LR test is based on a censored likelihood: the tail of the forecasted density is compared with the observed tail.

First, for different values of ρ the desired cutoff point VaR $= \Phi^{-1}(\rho)$ is computed. Then we define the new variable of interest as

$$z_t^* = \begin{cases} \text{VaR} & \text{if} \quad z_t \geq \text{VaR} \\ z_z & \text{if} \quad z_t < \text{VaR}. \end{cases}$$

The log-likelihood function for joint estimation of μ and σ^2 is

$$\begin{aligned} L(\mu, \sigma | z^*) &= \sum_{z^* < VaR} \log \frac{1}{\sigma} \phi \left(\frac{z_t^* - \mu}{\sigma} \right) + \sum_{z^* = VaR} \log \left(1 - \Phi \left(\frac{VaR - \mu}{\sigma} \right) \right) \\ &= \sum_{z^* < VaR} \left(-\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma} (z_t^* - \mu)^2 \right) + \sum_{z^* = VaR} \log \left(1 - \Phi \left(\frac{VaR - \mu}{\sigma} \right) \right). \end{aligned}$$

To construct the LR test the null hypothesis requires that $\mu = 0$, $\sigma^2 = 1$. Therefore the restricted likelihood L(0,1) is compared to the unrestricted one, L$(\hat{\mu}, \hat{\sigma}^2)$. The test statistic is then

$$LR_{\text{tail}} = -2(L(0,1) - L(\hat{\mu}, \hat{\sigma}^2)) \tag{2.5.8}$$

Under the null hypothesis, the test statistic is distributed $\chi^2(2)$.

Table 2.10 reports, for the different models considered and different assumptions for the residuals, the percentage of violations along with the p-value of the Berkowitz's test.

The relative number of violations is close to the theoretical one and assuming a t distribution for the residuals does not really improve the forecasting performances. For all the proposed specifications of the WAR model, the Berkowitz test does not reject the null hypothesis of appropriateness of the forecasted densities. Therefore all the models provide acceptable VaR forecasts. For the 1% VaR level, the results are somewhat surprising. The percentage of VaR violations is, for all the specifications, around 2.4% in front of a theoretical value of 1%. However, the p-values of the Berkowitz test are all higher than the rejection threshold of, say, 5%. This might be explained by the fact that the test proposed by Berkowitz is not a pointwise evaluation of the VaR violations, but rather analyzes the entire forecasted densities, or, in our case, the left tail of the distribution.

Besides the good forecasting performances of the proposed models, we want to stress the fact that there is no notable difference in the forecasting ability of the different specifications. Therefore, a very parsimonious (and thus quick to estimate) model like the restricted diagonal WAR is sufficient to model the riskiness of our portfolio.

2.6 Conclusions and direction for future research

In this paper we proposed a particular set of restricted specification of the WAR model for realized (co)variances. Our specifications rely on the ability to group assets according to some criterion, for example the economic sector, a common feature in the variance-covariance dynamics, and so on. This allowed us to drastically reduce the number of parameters. A comparison between the different specifications highlighted that there is no loss when a more parsimonious model is chosen. This is essentially due to the fact that the restricted model was justified by the data.

However, some aspects of the WAR process need to be clarified. In particular, the degrees of freedom seem to vary through time and it is not clear by which variables they are driven.

A straightforward extension of the present work involves applying the WAR model to solve concrete financial problems like dynamic portfolio choice, for instance.

This and other applications of the WAR model are left for future research.

Table 2.10: VaR failure rate and Berkowitz (2001) test's p-value

		10%	5%	1%
full WAR(1)	N	0.1072	0.0490	0.0230
		(0.6608)	(0.7038)	(0.8174)
	t	0.1041	0.0490	0.0230
		(0.8137)	(0.8508)	(0.9446)
block diagonal WAR(1)	N	0.1041	0.0521	0.0245
		(0.6441)	(0.6865)	(0.7984)
	t	0.1026	0.0505	0.0245
		(0.7836)	(0.8209)	(0.9157)
restr. block diag. WAR(1)	N	0.1057	0.0536	0.0245
		(0.6677)	(0.7093)	(0.8184)
	t	0.1041	0.0521	0.0245
		(0.7991)	(0.8341)	(0.9220)
diagonal WAR(1)	N	0.1057	0.0521	0.0245
		(0.6705)	(0.7121)	(0.8208)
	t	0.1041	0.0505	0.0245
		(0.7988)	(0.8337)	(0.9214)
restr. diag. WAR(1)	N	0.1057	0.0521	0.0245
		(0.6664)	(0.7080)	(0.8168)
	t	0.1041	0.0505	0.0245
		(0.7980)	(0.8329)	(0.9208)
full HAR-WAR	N	0.1103	0.0658	0.0291
		(0.0697)	(0.0800)	(0.1112)
	t	0.1087	0.0658	0.0260
		(0.1393)	(0.1574)	(0.2104)
block diag. HAR-WAR	N	0.1133	0.0536	0.0260
		(0.2612)	(0.2898)	(0.3711)
	t	0.1133	0.0536	0.0245
		(0.3929)	(0.4292)	(0.5297)
restr. block diag. HAR-WAR	N	0.1149	0.0551	0.0245
		(0.3722)	(0.4076)	(0.5057)
	t	0.1149	0.0551	0.0245
		(0.4991)	(0.5392)	(0.6474)
diagonal HAR-WAR	N	0.1118	0.0475	0.0245
		(0.4440)	(0.4831)	(0.5909)
	t	0.1103	0.0475	0.0245
		(0.5716)	(0.6141)	(0.7281)
restr. diag. HAR-WAR	N	0.1133	0.0475	0.0245
		(0.3707)	(0.4065)	(0.5063)
	t	0.1133	0.0475	0.0245
		(0.5333)	(0.5751)	(0.6881)

Appendix 2.A Relation between Wishart and gamma distribution

This proof follows the one in the Technical Appendix in Meucci (2005).

If $\mathbf{Y}$ is a Wishart distribution, then for any comfortable matrix $\mathbf{A}$ we have

$$\begin{aligned}\mathbf{AYA'} &= \mathbf{AX_1X_1'A'} + \cdots + \mathbf{AX_KX_K'A'} && (2.A.1)\\ &= \mathbf{Z_1Z_1'} + \cdots + \mathbf{Z_KZ_K'} && (2.A.2)\\ &\sim \mathrm{W}(K, \mathbf{A\Sigma A'}) && (2.A.3)\end{aligned}$$

since

$$\mathbf{X_t} \sim \mathrm{N}(0, \Sigma) \tag{2.A.4}$$

and

$$\mathbf{Z_t} \equiv \mathbf{AX_t} \sim \mathrm{N}(0, \mathbf{A\Sigma A'}). \tag{2.A.5}$$

By taking a row vector, i.e. $\mathbf{A} \equiv a'$, each term in the sum is normally distributed as follows:

$$Z_t \equiv a'\mathbf{X_t} \sim \mathrm{N}(0, a'\mathbf{\Sigma}a). \tag{2.A.6}$$

Now, for any random variable

$$y_i \sim \mathrm{N}(0, \sigma^2) \tag{2.A.7}$$

the gamma distribution with K degrees of freedom is defined as the distribution of the following variable:

$$x = y_1^2 + \cdots + y_K^2 \sim \mathrm{Ga}(K/2, 2\sigma^2). \tag{2.A.8}$$

and has p.d.f. of the form[11]

$$f(x|K/2, 2\sigma^2) = \frac{1}{(2\sigma^2)^{K/2}\Gamma(K/2)} x^{K/2-1} e^{x/2\sigma^2}. \tag{2.A.9}$$

Therefore from (3.3.14)

$$a'\mathbf{Y}a \sim \mathrm{Ga}(K/2, 2(a'\mathbf{\Sigma}a)). \tag{2.A.10}$$

Note that in Meucci (2005) we have $a'\mathbf{Y}a \sim \mathrm{Ga}(K, (a'\mathbf{\Sigma}a))$, because a different parametrization of the gamma distribution is used.

[11] Recall that if $x \sim Ga(a,b)$, then $f(x|a,b) = \frac{1}{b^a\Gamma(a)} x^{a-1} e^{x/b}$

Appendix 2.B Estimation of the degrees of freedom for a general WAR(p) process

We present here a way to derive the estimator of the degrees of freedom K in a general WAR(p) process. Differently from Chiriac (2007), we do not rely on the interpretation of a WAR process in terms of a Gaussian VAR process; in fact, for a WAR(p) process with $p > 1$ this interpretation is no longer valid (see Gourieroux et al., 2009). Instead, we use the fact that any portfolio of Wishart-distributed matrices follows a gamma distribution, as shown in the previous section.

Let $Y_t \in \mathbb{R}^n \times \mathbb{R}^n$ be a WAR(p) process:

$$E\left[Y_t | \mathbb{I}_{t-1}\right] = \sum_{j=1}^{p} M_j Y_{t-j} M_j' + K\Sigma. \tag{2.B.1}$$

where $\mathbb{I}_{t-1}$ is the information set available up to time $t-1$.

Under stationary conditions, the unconditional mean of the process, $E\left[Y_t\right]$, is obtained using the law of iterated expected values:

$$E\left[Y_t\right] = E\left[E\left[Y_t | \mathbb{I}_{t-1}\right]\right] = \sum_{j=1}^{p} M_j E\left[Y_{t-j}\right] M_j' + K\Sigma \tag{2.B.2}$$

As the unconditional distribution of any WAR(p) process is a centered Wishart distribution, applying the definition of centered Wishart distribution, we can write:

$$Y_t = \sum_{k=1}^{K} z_{k,t} z_{k,t}', \tag{2.B.3}$$

where $z_{t,k} \overset{i.i.d}{\sim} N(0, \Sigma(\infty))$.

From (3.B.3) we have that

$$\begin{aligned} E\left[Y_t\right] &= \sum_{k=1}^{K} E\left[z_{k,t} z_{k,t}'\right] \\ &= KV\left[z_{k,t}\right] \\ &= K\Sigma(\infty). \end{aligned} \tag{2.B.4}$$

Combining this result with (3.B.3) and defining $\Sigma^*(\infty) = K\Sigma(\infty)$ and $\Sigma^* = K\Sigma$ we get

$$\Sigma^*(\infty) = \sum_{j=1}^{p} M_j \Sigma^*(\infty) M_j' + \Sigma^* \tag{2.B.5}$$

From (3.3.14) we know that, for any given vector $\omega \in \mathbb{R}^n$

$$\omega' Y_t \omega \sim Ga(K/2, 2\omega'\Sigma(\infty)\omega). \tag{2.B.6}$$

Knowing the variance of a gamma-distributed random variable, we have

$$V\left[\omega' Y_t \omega\right] = \frac{K}{2}(2\omega'\Sigma(\infty)\omega)^2. \tag{2.B.7}$$

$\Sigma(\infty)$ is not observable, but given the estimated matrices $\hat{M}_j,\ j = 1, \dots, p$ and $\hat{\Sigma}^*$ we can recover $\hat{\Sigma}^*(\infty)$ that satisfies (3.B.5). Thus:

$$\begin{aligned} V\left[\omega' Y_t \omega\right] &= \frac{K}{2}\left(2\omega' \frac{\hat{\Sigma}^*(\infty)}{K}\omega\right)^2 && (2.B.8) \\ &= \frac{2}{K}\left(\omega'\hat{\Sigma}^*(\infty)\omega\right)^2. && (2.B.9) \end{aligned}$$

Therefore the estimated degrees of freedom are

$$\hat{K} = \frac{2(\omega'\hat{\Sigma}^*(\infty)\omega)^2}{V\left[\omega' Y_t \omega\right]} \tag{2.B.10}$$

Manuscript 3

Estimating the degrees of freedom of the Realized Volatility Wishart Autoregressive model

3.1 Introduction

Risk management, option pricing and asset allocation are, among others, three fields in finance whose key element is the ability to model, estimate and predict volatility. The increased availability of high-frequency data provides new tools for forecasting variances and covariances between assets. In particular, after the seminal paper of Andersen and Bollerslev (1998), the literature on realized volatility has grown enormously; see McAleer and Medeiros (2006) for a review. While most works focus on the study of univariate series, recently there has been growing theoretical and empirical interest in extending the results for the univariate process to a multivariate framework. In this context, two pioneering contributions have been made by Barndorff-Nielsen and Shephard (2004) and Bandi and Russel (2005). Barndorff-Nielsen and Shephard (2004) did not consider the presence of microstructure noise, whereas the noise has been considered in Bandi and Russel (2005).

Alternative approaches to the high-frequency covariance estimator have recently been introduced by Hayashi and Yoshida (2005, 2006), Sheppard (2006) and Zhang (2006), among others. For example, instead of using calendar returns, the Hayashi and Yoshida estimator (HY) is based on overlapping tick-by-tick returns. Sheppard (2006) analyzed the conditions under which the realized covariance is an unbiased and consistent estimator of the integrated covariance. Zhang (2006) also studied the effects of microstructure noise and non-synchronous trading in the estimation of integrated covariance between assets.

Although the literature on multivariate extensions of the realized variance regarding the definition of new estimators of the realized covariances resulted in a notable amount of academic works, only a few papers provide financial applications for these new estimators.

One explanation for the scarcity of empirical contributions in multivariate realized volatility analysis is the difficulty in finding a dynamic specification of a stochastic volatility matrix which satisfies the symmetry and positivity properties of each forecasted matrix, does not suffer from the so called 'curse of dimensionality' and possesses a closed-form expression for the forecasts at any horizon.

A solution to this problem is represented by the Wishart autoregressive model (WAR) proposed by Gourieroux et al. (2009). The model is based on a dynamic extension of the Wishart distribution, a family of probability distributions for nonnegative-definite matrix-valued random variables. This specification is compatible with financial theory, satisfies the constraints on volatility matrices, has a flexible form and, most importantly, maintains the coefficients' interpretability. A specific parametrization of the WAR model has recently been introduced in Bonato et al. (2008). In particular, they show how to achieve a great reduction of the number of parameters according to an economic criterion which is consistent with standard sectorial asset allocation approaches. The parametric structure they propose imposes a block structure on the coefficient matrices, and the model is named *block WAR*. For a detailed discussion on the linear and non-linear causality restrictions see Gourieroux and Sufana (2007) or Jasiak and Lu (2007).

One drawback of the WAR model is the possibility that it is degenerate, i.e. the forecasted realized covariance matrix is not assured to be positive semi-definite, or worse, that the density is not defined. As illustrated in the next session, for any positive integer K, any Wishart process of dimension n, with K degrees of freedom and innovation matrix Σ, can be interpreted as the sum of the cross product of K independent normally distributed random vectors, with zero mean and common covariance matrix Σ. The process is non degenerate if $K > n$ and the condition that assures the existence of the density is $K > n - 1$. It is immediate to see that, in empirical application, problems arise when the value of the estimated degrees of freedom is low and consequently these conditions do not hold. The search for a valid estimator of K is therefore crucial and important is also to understand in which cases the WAR model is not appropriate for modeling the sequence of variance-covariance matrices of a group of assets.

A first work in this direction was made by Chiriac (2006, 2007). These papers empirically analyzed the properties of the WAR model applied to a series of intraday realized volatility matrices. It is shown that large variations in time and high persistence in the volatility estimators cause misidentification of the model. In Chiriac (2007), a modified representation of the WAR model is proposed. This new representation is able to capture the dynamics if the variance-covariance matrices with large variation in time. An extended application of the model, originally defined under stationary assumptions, indicates that multivariate volatility processes with large variance follow a degenerate Wishart distribution.

Within this paper we present an in-depth study of the estimation of the WAR model and focus on three main points. First, we introduce an alternative estimator for the degrees of freedom. Differently from the standard estimator presented in Gourieroux et al. (2009), and used in Chiriac(2006, 2007), this novel estimator is not function of other parameters of the model (see next section for the details) but relies on the fact that the distribution of the volatility of any portfolio is, under the WAR assumption, gamma distributed with shape parameter $K/2$, where K denotes the degrees of freedom. We claim that this alternative estimator works better than the standard one as, in this latter, the estimation error of the other parameters of the model is carried inside the estimate of K. A simulation experiment confirms our conjecture. Second, we study one possible cause for low values of the estimated degrees of freedom. We claim this is related to the definition of Wishart distribution. It comes in fact, from cross product of Gaussian random vectors. The normal distribution is well known to have thin tails and thus is not able to capture extreme market events. We think that this 'thin tail' aspect is brought inside the Wishart model, which is thus incapable to capture extreme events in the variance-covariance process. A simulation study supports our hypothesis and also shows that the alternative estimator is less affected by the presence of extreme events in the variance process. To capture extreme events a mixture model is proposed. In this model a small weight is associated with a component having a strong variance. A closed-form expression is derived for the method of moments estimator of K. It shows that any perturbation to the process induces a bias toward zero of the estimated degrees of freedom, confirming our previous findings. Third, we empirically study the two estimators when the WAR model is applied to the sequence of realized covariance matrices. We use three estimators of the realized covariance matrix at different sampling frequencies. We find that the degrees of freedom increase with the sampling frequency and do not appear to be function of the variance of the variance-covariances estimators. We conclude the paper noting that the degrees of freedom are likely to be varying over time and, in our opinion, the best strategy is to use a rolling window to estimate the model. This is in line, for example, with the way banks and regulators implement risk management measures.

The paper is organized as follows. Section 3.2 introduce the WAR model for multivariate realized volatility of Gourieroux et al. (2009). Section 3.3 presents the alternative way of estimating the degrees of freedom of the WAR model. The estimation of K in a misspecified WAR model and when outliers are introduced in the process is studied in Section 3.4. In Section 3.5 the mixture of WAR model is presented and analyzed. An empirical application of the two estimators using high-frequency data is showed in Section 3.6 whereas consequences of the WAR under cointegration, as proposed in Chiriac (2007) are analyzed in Section 3.7. Section 3.8 concludes and gives directions for future research.

3.2 The Wishart Autoregressive Model

This section gives a brief illustration of the Wishart Autoregressive Model. We will follow, in the exposition, Gourieroux et al. (2009). Refer to the same paper for a more detailed presentation on the WAR model.

The Wishart process is a process (Y_t) formed by stochastic symmetric positive definite matrices of dimension $n \times n$. Let us consider the process Y_t defined by

$$Y_t = \sum_{k=1}^{K} x_{k,t} x'_{k,t}, \tag{3.2.1}$$

where $x_{k,t}, k = 1, \ldots, K$ are independent Gaussian VAR(1) processes of dimension n with the same autoregressive parameter matrix M and innovation variance Σ:

$$x_{k,t} = M x_{k,t-1} + \epsilon_{k,t}, \qquad \epsilon_{k,t} \sim N(0, \Sigma). \tag{3.2.2}$$

The process Y_t is called Wishart Autoregressive Process of order 1, WAR(1), denoted by $\mathrm{W}(K, \Sigma, M)$. The transition density of WAR(1) depends on the following parameters: K is the scalar degree of freedom, strictly greater than $n-1$, M is the $n \times n$ matrix of autoregressive parameters, and Σ is an $n \times n$ symmetric, positive definite matrix. The process defined in (3.2.2) is (strictly) stationary, i.e. the process defined in Equation (3.2.1) is (strictly) stationary, if and only if the matrix M has root with modulus (strictly) less than 1. If the eigenvalues of M are all real, nonnegative and strictly less than one, this indicates that the process can be considered as a time discretized version of a continuous time process[1], and satisfies the stationarity conditions. An important property of the Wishart distribution is that the matrices Y_t are positive definite if and only if $K \geq n$ and for a non-centered Wishart specification, the distribution of Y_t possesses a density function only when $K > n-1$ (hence the condition above). Thus, for $K < n-1$ no density can be defined and for $K < n$ the process Y_t is given by a sequence of singular covariance matrices with degenerate Wishart distribution (Muirhead, 1982). We stress that the interpretation of Y_t from latent Gaussian VAR(1) processes is valid for integer valued K only and, in general, any economic or financial interpretation of the latent processes $(x_{k,t})$ is not necessary. The dynamic of a Wishart autoregressive process for any $K > n-1$ is specified by its conditional Laplace transform, which defines the conditional expectations of any exponential transformation of element of the matrix Y_{t+1} (see Gourieroux et al., 2009 for more details):

$$\begin{aligned}\mathbf{\Psi}_t(\Gamma) &= E[\exp Tr(\Gamma Y_{t+1})] \\ &= \frac{\exp Tr\left[M'\Gamma(\mathbf{I}_d - 2\Sigma\Gamma)^{-1}MY_t\right]}{[\det(\mathbf{I}_d - 2\Sigma\Gamma)]^{K/2}}.\end{aligned} \tag{3.2.3}$$

From Proposition 2 in Gourieroux et al. (2009) we have:

$$E_t\left(Y_{t+1}\right) = MY_tM' + K\Sigma. \tag{3.2.4}$$

The first conditional moment is affine function of the lagged values of the volatility process. In particular, the WAR(1) process is a weak linear AR(1) process. More precisely we get:

$$Y_{t+1} = MY_tM' + K\Sigma + \eta_{t+1}, \tag{3.2.5}$$

where η_{t+1} is a matrix of stochastic errors with conditional mean zero. Equivalently, we get:

$$vech(Y_{t+1}) = A(M)vech(Y_t) + vech(K\Sigma) + vech(\eta_{t+1}), \tag{3.2.6}$$

where $vech(Y)$ denotes the vector obtained by stacking the lower triangular elements of Y and $A(M)$ is a function of M. The error term η is a weak white noise, since it features conditional heteroskedasticity and, even after conditional standardization, is not identically distributed.

3.2.1 Estimation of M and Σ

Three 'objects' need to be estimated to get all the parameters necessary to perform predictions of the WAR model: the latent autoregressive matrix M, the innovation covariance matrix Σ and the degrees of freedom K.

[1] This can be useful in further financial application, like derivative pricing in continuous time. See e.g. Gourieroux and Sufana (2003, 2004), Gourieroux et al. (2004).

Under the assumption that $K > n-1$ it is straightforward to show that:

i) K and Σ are identifiable while the autoregressive coefficients in M are identifiable up to their sign.

ii) Σ is first order-identifiable up to a scale factor and M is first-order identifiable up to its sign. The degrees of freedom K are not first-order identifiable but are second-order identifiable.

The estimation of K will be defined in detail in the next section, whereas the estimation of M and Σ is presented here, either for the general case and for the diagonal M case. The assumption of M diagonal was first introduced within the WAR context in Bonato et al. (2008) to reduce the number of parameters to be estimated. From an economic point of view, it assumes no spillover between variances.

Following Gourieroux et al. (2009), the first order conditional moments can be used to calibrate the parameters in M and Σ, up to the sign and scale factor, respectively.

As the first-order method of moments is equivalent to nonlinear least squares, the estimator is defined as:

$$\left(\hat{M}, \hat{\Sigma}^*\right) = \mathrm{Argmin}_{M,\Sigma^*} S^2\left(M, \Sigma^*\right)$$

where

$$\begin{aligned} S^2\left(M, \Sigma^*\right) &= \sum_{t=2}^{T} \sum_{i<j} \left(Y_{ij,t} - \sum_{k=1}^{n} \sum_{l=1}^{n} Y_{kl,t-1} m_{ik} m_{lk} - \sigma_{ij}^*\right)^2 \\ &= \sum_{t=2}^{T} \| vech(Y_t) - vech(M Y_{t-1} M' + \Sigma^*) \|^2 \end{aligned}$$

and $\Sigma^* = K\Sigma$.

As mentioned in Gourieroux et al. (2009), any statistical software which accounts for heteroskedasticity can be used to the get the estimates. We present here the complete procedure under the assumption that M is diagonal as we want to emphasize the quickness of the algorithm.

For each $Y_t, t = 1, \ldots, T$ of dimensions $n \times n$, we consider the matrix $\mathbf{Y}$, of dimensions $T \times n(n+1)/2$ build with the *vech* of Y_t for each time $t = 1, \ldots, T$; i.e. the i-th row of $\mathbf{Y}$ is $vech(Y_i)$.

Under the hypothesis that M is diagonal, define $a = diag(M)$ and $dg(a)$ the diagonal matrix with the vector a as diagonal. Then

$$M Y_{t-1} M' = dg(a) Y_{t-1} dg(a) = (aa') \odot Y_{t-1} \tag{3.2.7}$$

and

$$vech(M Y_{t-1} M') = vech(aa') \odot vech(Y_{t-1}), \tag{3.2.8}$$

where $\odot$ denotes the elementwise product. Define $[\mathbf{Y}]_2^T$ as the matrix obtained from $\mathbf{Y}$ when dropping the last row, i.e. considering the time from T down to time 2. Define $A = vech(aa')$ and $Z = vech(\Sigma^*)$. The residual matrix W is obtained as

$$W = [\mathbf{Y}]_2^T - \left(A' \otimes i_{T-1}\right) \odot [\mathbf{Y}]_1^{T-1} - Z' \otimes \mathbf{i}_{T-1} \tag{3.2.9}$$

where $\mathbf{i}_{T-1}$ is a $T-1 \times 1$ vector of ones $\otimes$ denotes the Kronecher product.

Then the minimization problem reduces to:

$$\left(\hat{M}, \hat{\Sigma}^*\right) = Arg \min_{M,\Sigma^*} \left[\mathbf{i}'_{T-1} \left(W \odot W\right) \mathbf{i}_{n(n+1)/2}\right]. \tag{3.2.10}$$

3.3 Estimation of the degrees of freedom

Whereas the estimation of the entries of the autoregressive matrix M and of the innovation variance Σ (up to multiplication for a scale parameter) is relatively straightforward, the estimation of the degrees of freedom poses some challenges. We first present the estimation procedure introduced in Gourieroux et al. (2009) and then show how the same parameter K can be estimated relying on the fact that, given a portfolio allocation α, its volatility $\alpha' Y_t \alpha$ is gamma-distributed with a shape parameter equal to $K/2$.

Consider the simple WAR(1) model. The marginal distribution of the WAR(1) is the centered Wishart distribution, defined $W(K, 0, \Sigma(\infty))$, where $\Sigma(\infty)$ is computed from

$$\Sigma(\infty) = M\Sigma(\infty)M' + \Sigma. \tag{3.3.1}$$

Thus, the unconditional variance of a portfolio volatility is given by:

$$V(\alpha' Y_t \alpha) = \frac{2}{K}[\alpha'\Sigma^*(\infty)\alpha]^2, \tag{3.3.2}$$

where α is a vector of dimension $(n \times 1)$ and $\Sigma^*(\infty) = K\Sigma(\infty)$. A consistent estimator of the degrees of freedom K can be computed as follows:

Step 1 Compute $\hat{\Sigma}^*(\infty)$ as solution of

$$\hat{\Sigma}^*(\infty) = \hat{M}\hat{\Sigma}^*(\infty)\hat{M}' + \hat{\Sigma}^*. \tag{3.3.3}$$

Step 2: Chose a portfolio allocation and compute its sample volatility

$$V(\alpha' Y_t \alpha) = \frac{1}{T}\sum_{t=1}^{T}\left[\alpha' Y_t \alpha - \frac{1}{T}\sum_{t=1}^{T}\alpha' Y_t \alpha\right]^2. \tag{3.3.4}$$

Step 3: A consistent estimator of K is:

$$\hat{K}(\alpha) = 2[\alpha'\hat{\Sigma}^*(\infty)\alpha]^2/\hat{V}(\alpha' Y_t \alpha) \tag{3.3.5}$$

Step 4: A consistent estimator of Σ is $\hat{\Sigma}(\alpha) = \hat{\Sigma}^*/\hat{K}(\alpha)$.

This method provides consistent estimates of the degrees of freedom but is problematic in two aspects: first, it needs to estimate the matrix $\Sigma(\infty)$; second, it makes use of the estimates $\hat{M}$ and $\hat{\Sigma}$, drawing their estimation error in the estimate of $\hat{K}$.

A more direct way that does not need to rely on the estimates of M and Σ comes from the distribution of the volatility of a portfolio under the Wishart assumption. On this alternative estimation procedure will focus this section.

3.3.1 Introducing an alternative estimator

Consider a portfolio allocation $\alpha \in \mathbb{R}^n$. We know that the unconditional distribution of Y_t is $W(K, 0, \Sigma(\infty))$, a centered Wishart distribution. The following theorem gives the distribution of the volatility of any portfolio with allocation α.

Theorem 1. *Let Y_t, $t = 1, \ldots, T$ a sequence of $n \times n$ matrices from a $W[K, M, \Sigma]$ process. Then, for any vector $\alpha \in \mathbb{R}^n$ we have that:*

$$\alpha' Y_t \alpha \sim Ga\left(\frac{K}{2}, 2\alpha'\Sigma(\infty)\alpha\right), \tag{3.3.6}$$

Proof: This proof follows the one of Meucci (2005) in the Technical Appendix, p. 33-34 of the book.

If $\mathbf{Y}$ is a Wishart distribution, then for any comfortable matrix $\mathbf{A}$ we have

$$\begin{aligned} \mathbf{AYA'} &= \mathbf{AX_1X_1'A'} + \cdots + \mathbf{AX_KX_K'A'} && (3.3.7)\\ &= \mathbf{Z_1Z_1'} + \cdots + \mathbf{Z_KZ_K'} && (3.3.8)\\ &\sim \mathrm{W}(K, \mathbf{A\Sigma A'}) && (3.3.9) \end{aligned}$$

since

$$\mathbf{X_t} \sim \mathrm{N}(0, \Sigma) \tag{3.3.10}$$

and

$$\mathbf{Z_t} \equiv \mathbf{AX_t} \sim \mathrm{N}(0, \mathbf{A\Sigma A'}). \tag{3.3.11}$$

By taking a row vector, i.e. $\mathbf{A} \equiv a'$, each term in the sum is normally distributed as follows:

$$Z_t \equiv a'\mathbf{X_t} \sim \mathrm{N}(0, a'\mathbf{\Sigma}a). \tag{3.3.12}$$

Now, for any random variable

$$Y_i \sim \mathrm{N}(0, \sigma^2) \tag{3.3.13}$$

the gamma distribution with K degrees of freedom is defined as the distribution of the following variable:

$$Y_1^2 + \cdots + Y_K^2 \sim \mathrm{Ga}(K/2, 2\sigma^2), \tag{3.3.14}$$

and has p.d.f. of the form[2]

$$f(x|K/2, 2\sigma^2) = \frac{1}{(2\sigma^2)^{K/2}\Gamma(K/2)} x^{K/2-1} e^{x/2\sigma^2}. \tag{3.3.15}$$

Therefore from (3.3.14)

$$a'\mathbf{Y}a \sim \mathrm{Ga}(K/2, 2(a'\mathbf{\Sigma}a)). \tag{3.3.16}$$

Note that in Meucci (2005) we have $a'\mathbf{Y}a \sim \mathrm{Ga}(K, (a'\mathbf{\Sigma}a))$, because a different parametrization of the gamma distribution is used. ■

Theorem 1 states that the distribution of any portfolio with allocations α follows a gamma distribution with the degrees of freedom K as shape parameter. Then, an unbiased estimator of K is be obtained simply via maximum likelihood (ML) by fitting a gamma distribution to the process $\alpha'\Sigma(\infty)\alpha$[3]. In this way, the estimator is only function of the portfolio allocation α and does not embody the estimation error of $\hat{M}$ and $\hat{\Sigma}^*(\text{inf})$. Recall also that the ML estimator is, among all the possible unbiased estimators, the more efficient. For these reasons we expect this novel estimator to be also more efficient compared with the standard one. In the sequel we will denote with K_G and K_B the estimators as presented in Gourieroux et al. (2009) and in this paper, respectively. A comparison of the efficiency of two estimators is presented in the next subsection.

Theorem 1 turns out to be useful also to prove that the first estimation procedure presented can be extended to a more general WAR(p), $p \geq 1$ setting. The proof given for example in Chiriac (2007) (see Appendix 3.A) relies on the fact that a WAR(1) process can be interpreted as the sum of squares of autoregressive Gaussian processes. For a WAR(p) process this interpretation is no longer valid. A proof for a general WAR(p) is reported in Appendix .

3.3.2 Comparison of the two estimators

As shown in Chiriac (2007), the first estimator proposed, K_G, has the following asymptotic distribution:

$$\sqrt{T}(\hat{K}_G - K) \overset{d}{\sim} N(0, \sum_{j=-\infty}^{\infty} \gamma_j) \tag{3.3.17}$$

[2]Recall that if $x \sim Ga(a,b)$, then $f(x|a,b) = \frac{1}{b^a\Gamma(a)} x^{a-1} e^{x/b}$

[3]When performing the ML estimation one should be careful to the parametrization of the gamma density function. According to Meucci's notation, it would be for instance $\alpha' Y_t \alpha \sim \mathrm{Ga}(K, \alpha'\Sigma(\infty)\alpha)$

where $\gamma_j = E[(S_t - \mu)(S_{t-j} - \mu)]$ with $S_t = \sum_{k=1}^{K} \frac{(\alpha' x_{k,t})^2}{\alpha' \Sigma(**) \alpha}$. $\mu \equiv E[S_t] = K$, α and $x_{k,t}$ are as defined previously and $\Sigma(**)$ is such that

$$\Sigma(**) = \Sigma + M\Sigma M' + M^2\Sigma(M^2)' + \cdots + M^{T-1}\Sigma(M^{T-1})'.$$

The second estimator, if maximum likelihood is used to estimates the parameters is distributed as

$$\sqrt{T}(\hat{K}_B - K) \overset{asy}{\sim} N(0, \frac{2K}{\psi'(\frac{K}{2})\frac{K}{2} - 1}) \qquad (3.3.18)$$

where $\psi(\eta) = d \log \Gamma(\eta)/d\eta$ is the digamma function.

An analytical comparison of the variance of the two estimators is not feasible and therefore we rely on simulations to test their efficiency.

Considered a diagonal autoregressive matrix M and a covariance matrix Σ:

$$M = \begin{pmatrix} 0.8 & 0 & 0 \\ 0 & 0.2 & 0 \\ 0 & 0 & 0.35 \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

We simulated 3×3 matrices coming from a WAR distribution with 7 degrees of freedom and sample length 20, 250, 500 and 1000. Then we estimated the degrees of freedom. We repeated this operation 1000 times. Mean, standard deviation, mean square error (MSE), mean absolute error (MAD), skewness, kurtosis and Bera and Jarque (1980) normality test (JB test hereafter) p-values and are reported in Table 3.1. Both the estimators are unbiased, except when the sample size is very small. We also see that, independently of the sample length, K_B always has a lower standard deviation, MSE and MAD with respect to its competitor. K_G has a distribution closer to the normal (except for T=25), as indicated by the JB test when the sample length is 1,000. Figure 3.1 plots the kernel distribution of the two estimators. They are clearly both unbiased but the density of $\hat{K}_B$ has a smaller dispersion around the mean, as confirmed by Table 3.1. To conclude, from a MSE and MAD perspective, K_B seems to be preferable to K_G for any sample size.

3.4 Misspecified Wishart Autoregressive model

The estimation of the degrees of freedom of a Wishart process is of fundamental importance to assess the appropriateness of the model. In fact, if K is smaller than n (n is the dimension of the process), the covariance process has a degenerate Wishart distributions. Moreover, if K is smaller than $n - 1$, no density function can be defined for the variance-covariance distribution. Therefore, in empirical situations where the degrees of freedom are found not to be large enough, the WAR model becomes useless.

Table 3.1: The sample mean, sample standard deviation, the mean of the squared error (MSE) and the mean of the absolute deviation (MAS), skewness, kurtosis and Jarque-Bera normality test p-value, for the estimators of the degrees of freedom K.

	Mean	S.d.	MSE	MAD	Skewness	Kurtosis	J-B test
				T = 25			
$\hat{K}_B$	8.7233	3.3558	14.2203	2.5703	1.4503	5.8371	0.0010
$\hat{K}_G$	8.7884	4.3816	22.3781	2.9568	5.2833	76.9502	0.0010
				T = 250			
$\hat{K}_B$	7.1039	0.6618	0.4483	0.5267	0.3268	3.2919	0.0010
$\hat{K}_G$	7.0939	0.7848	0.6241	0.6246	0.3095	3.1064	0.0017
				T = 500			
$\hat{K}_B$	7.0496	0.4769	0.2297	0.3759	0.3873	3.3472	0.0010
$\hat{K}_G$	7.0295	0.5520	0.3053	0.4361	0.3146	3.3727	0.0010
				T = 1000			
$\hat{K}_B$	7.0343	0.3394	0.1162	0.2703	0.1712	3.1977	0.0387
$\hat{K}_G$	7.0205	0.3975	0.1582	0.3124	0.1327	3.1711	0.1161

* 0.001 and 0.5 are the smallest and the biggest tabulated p-values for the JB test for small samples in the Matlab function `jbtest.m`

Chiriac (2006) studies the empirical properties of the WAR model and focuses in particular on the estimated degrees of freedom. She found that value of Wishart degrees of freedom, K, estimated from a sample larger than a month, or from data aggregated at lower level (30 minutes), indicates that the volatility process has a degenerate Wishart distribution. She concludes that the estimation of the Wishart process is favored by the time intervals with relatively small variance in the value of the volatility estimator. She also shows that the estimated degrees of freedom are inversely related to the variance of the volatility estimators. One possible explanation to these results is that, since an increase in the length of the series induces an increase in the variance of the volatility estimators, the performance of the WAR model declines when estimated over samples larger than one month.

As in Bonato et al. (2008) samples going from 1997 until 2005 are used to construct the series of realized covariances matrices and the process is found to be stationary and non-degenerate, we argue that the length of the series does not represent an issue regarding the stationarity and the estimation of the degrees of freedom. A more appropriate explanation is, in our opinion, related to an other matter: the presence of extreme observations in the process.

By definition, a WAR(1) process and in general any Wishart process, can be interpreted in terms of Gaussian vectors, assumed that the degrees of freedom K are integer. The assumption of Gaussian distribution for a time series has a first, fundamental, implications: thin tails. This means that extreme events are unlikely to happen and, in general, a normal distribution does

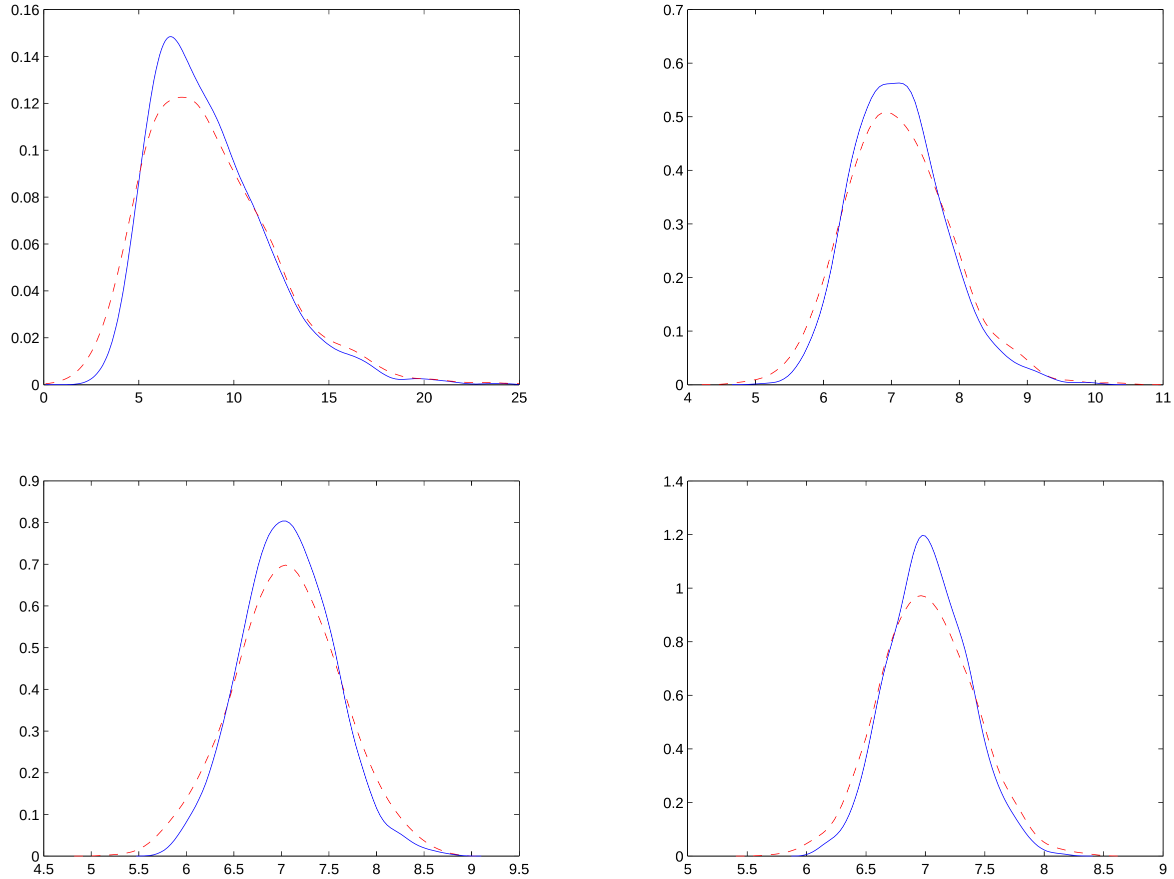


Figure 3.1: Kernel densities of $\hat{K}_B$ (blue solid line) and $\hat{K}_G$ (red dashed line) when T = 20 (top left), T = 250 (top right), T=500 (bottom left) and T=1,000 (bottom right).

not account for them. Our hypothesis about a degenerate WAR process is very simple. If, for an integer K, the series of variances-covariances can be interpreted as the sum of the cross-product of K gaussian random vectors, the 'thin tail' property of the normal distribution is somehow conveyed into the sequence of Wishart matrices. Roughly speaking, extreme events in the series of the variance-covariance matrices are not expected from this model. In particular, we do not claim that the WAR model is unappropriate when variances-covariances display large variation in time (this may be captured allowing the degrees of freedom to be time varying), but rather when extreme observations are present in the process.

Here is a simple example. Consider the WAR(1) process Y_t with $K = 7$ degrees of freedom, latent autoregressive matrix M and covariance matrix Σ

$$M = \begin{pmatrix} 0.4472 & 0 \\ 0 & 0.7071 \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} 0.8 & 0 \\ 0 & 0.5 \end{pmatrix}.$$

With M diagonal matrix, all the entries of each $x_{t,k} = [x_{k,t}(1) \quad x_{k,t}(2)]'$ follow a simple univariate AR(1) process where the autoregressive coefficients are the diagonal entries of M, i.e.

$$\begin{aligned} x_{k,t}(1) &= 0.4472 x_{k,t-1}(1) + \epsilon_{t,1} \\ x_{k,t}(2) &= 0.7071 x_{k,t-1}(2) + \epsilon_{t,2} \end{aligned}$$

where $\epsilon_{t,1} \sim N(0, 0.8)$ and $\epsilon_{t,2} \sim N(0, 0.5)$. The first entries of Y_t, say $Y_t(1,1)$ is the (realized) variance of the first asset at time t and it is given by:

$$Y_t(1,1) = \sum_{k=1}^{7} x_{k,t}(1)^2 \tag{3.4.1}$$

where $x_{k,t}(1)$ denotes the first element of the autoregressive gaussian vector $x_{k,t}$. The particular values of the variances of the error terms are chosen in such a way that the unconditional distribution of $x_{k,t}(1)$ and $x_{k,t}(2)$ is the standard normal.[4] This implies that each $x_{k,t}(1)^2$ is χ^2_1 distributed and thus $Y_t(1,1)$, being a sum of χ^2_1 random variables is χ^2_7 distributed. Therefore, the probability to observe a value for the realized variance of the first asset exceeding, say, 18 is less than 1%. The number we give, 18, has not a real meaning as, we repeat, it comes from the particular assumption on Σ and M, in this case set to be diagonal. However, it helps to show why extreme observations in the variance-covariance process are not expected in a Wishart model.

3.4.1 An example with real data

We present now a more realistic example. We use the tick-by-tick transaction prices on the S&P 500 and NASDAQ 100 futures recorded at the Chicago Mercantile Exchange[5] (CME) during its regular trading hours (RTH), i.e. from 8.30am to 3.15pm Central Time (CT) . The samples cover the period from March 3, 2003 to October 31, 2008 for a total of 1344 trading days. The original data are pre-processed to eliminate obvious data errors (transaction prices reported at zero, transaction time out of the order, etc...) and "bounce back" outliers larger than the cutoff 0.025. As stated in Aït-Sahalia and Mancini (2008), other cutoff thresholds could be conceived, but they would be all equally arbitrary. Different values were tried as threshold and 0.025 gave us the most satisfactory result.

We adopted the conventional[6] practice of using the future contract with the largest trading volume. As the contract approached maturity (usually one week before the maturity of that contract), we moved to the next contract, ensuring no overlapping periods in the price sequence and no returns computed on prices from different contracts. For the days in the sample period

[4] Recall that for a general AR(1) process $x_t = \rho x_{t-1} + \epsilon_t$, where $\epsilon_t \sim N(0, \sigma^2)$, the unconditional variance of the process is $\sigma^2/(1-\rho^2)$.

[5] The data were provided by www.opentick.com

[6] As done in Martens and van Dijk (2007) and de Pooter et al. (2006) among others.

this results in on average 2,190 transaction prices during the trading floor hours for the S&P 500 futures and 570 transaction prices for the NASDAQ 100 futures.

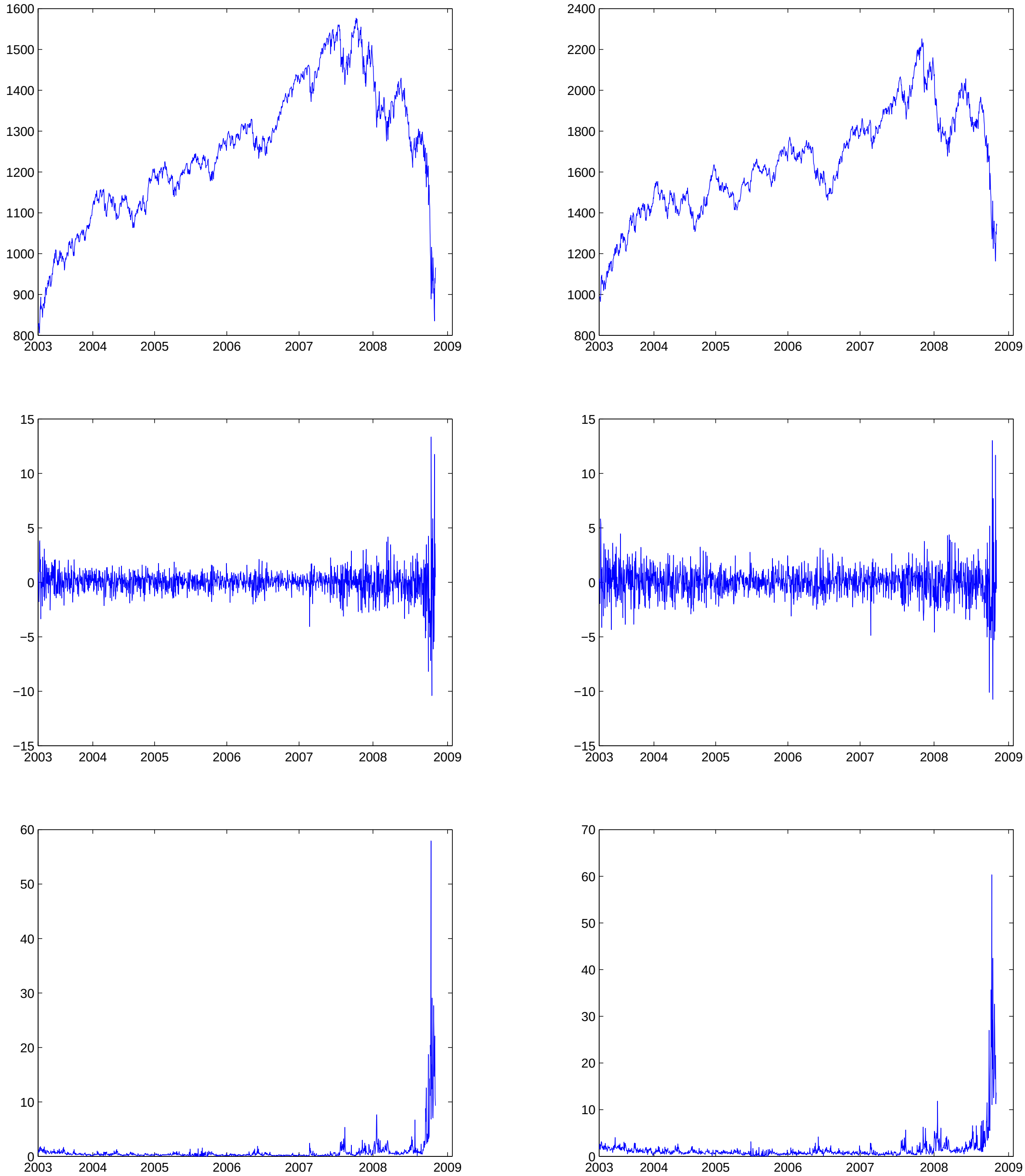


Figure 3.2: Daily prices (top), daily returns (middle) and TSRV (bottom) for the S&P 500 (left) and NASDAQ 100 (right) series.

To construct the series of realized volatility we adopted the two time scales estimator (TSRV) of Zhang et al. (2005). Consider, for simplicity, a trading day of length $h = 1$. Assume availability of $I + 1$ equally-spaced logarithmic asset prices over $[0, 1]$ and write

$$p_{j,\delta} = p^*_{j,\delta} + \eta_{j,\delta}$$

or, in terms of continuosly-compounded returns,

$$\underbrace{p_{j,\delta} - p_{(j-1),\delta}}_{r_{j,\delta}} = \underbrace{p^*_{j,\delta} - p^*_{(j-1),\delta}}_{r^*_{j,\delta}} + \underbrace{\eta_{j,\delta} - \eta_{(j-1),\delta}}_{\epsilon_{j,\delta}}$$

where p^* denotes the unobservable equilibrium price, η denotes the unobservable market microstructure noise and $\delta = 1/M$ is the time distance between adjacent observations.

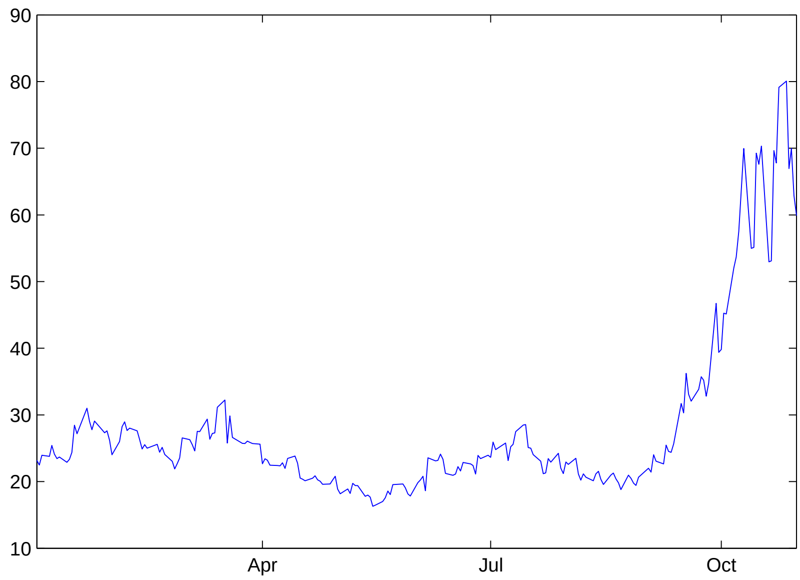


Figure 3.3: VIX index prices from 02 January to 31 October 2008. Source: finance.yahoo.com

Assume we divide the original grid of I arrivals in q non-overlapping subgrids $\Psi^{(i)}$, with $i = 1, \ldots, q$. The first sub-grid starts at t_0 and takes every q-th arrival time, i.e. $\Psi^{(1)} = (t_0, t_{0+q}, t_{0+2q}, \ldots)$, the second subgrid starts at t_1 and also takes every q-th arrival time, i.e. $\Psi^{(2)} = (t_1, t_{1+q}, t_{1+2q}, \ldots)$, and so on. Given the generic i-th subgrid of arrival times, the corresponding realized variance estimator is defined as

$$\widehat{rv}^{(i)} = \sum_{t_j, t_{j+} \in \Psi^{(1)}} (p_{t_{j+}} - p_{t_j})^2 \tag{3.4.2}$$

where t_j and t_{j+} denote the adjacent elements in $\Psi^{(i)}$.

Table 3.2: Descriptive statistics for daily log-returns, standardized log-returns and daily realized volatility for S&P 500 and NASDAQ 100 (4 March 2003 to 31 October 2008).

	Mean	S.d.	Skewness	Kurtosis	Max	Min	JB-test*
			r_t				
S&P500	-0.0048	1.0096	-0.7862	19.2757	7.4562	-9.2567	0.0010
NASDAQ	-0.0223	1.2719	-0.6168	11.8597	8.5143	-10.0122	0.0010
			$r_t/RV_{TS,t}^{1/2}$				
S&P500	0.0600	1.0408	-0.0616	2.7403	2.7720	-2.9548	0.1024
NASDAQ	0.0293	0.9556	-0.0237	2.8450	3.0077	-2.6032	0.4846
			$RV_{TS,t}^{1/2}$				
S&P500	0.7717	0.5936	5.4294	48.1408	8.8929	0.2104	0.0010
NASDAQ	1.1513	0.6500	4.6770	35.6557	8.6528	0.3706	0.0010
			$\log(RV_{TS,t}^{1/2})$				
S&P500	-0.4004	0.4748	1.1837	5.8926	2.1852	-1.5590	0.0010
NASDAQ	0.0522	0.3838	1.1959	6.4987	2.1579	-0.9928	0.0010

* 0.001 and 0.5 are the smallest and the biggest tabulated p-values for the JB test for small samples in the Matlab function `jbtest.m`

The two time scales estimator of realized volatility is constructed as

$$\widehat{RV}_{TS} = \frac{\sum_{i=1}^{q} \widehat{rv}^{(i)}}{q} - \overline{I}\widehat{\mathbb{E}}(\epsilon^2), \tag{3.4.3}$$

where $\overline{I} = \frac{I-q+1}{q}$, $\widehat{\mathbb{E}}(\epsilon^2) = \frac{\sum_{j=1}^{I}(p_{t_{j+}} - p_{t_j})^2}{I}$ is a consistent estimator of the second moment of the noise return and $\bar{I}\hat{\mathbb{E}}(\epsilon^2)$ is a bias-correction. This estimator averages the realized variance estimates obtained with subsampling and bias-corrected them estimating the bias due to the noise using all the observations. To compute the TSRV we used, as done in Brownlees and Gallo (2008), a fixed sampling frequency equal to 15 seconds to estimate the bias due to the noise and, as done in Aït-Sahalia and Mancini (2008), a slow time scale of five minutes. Figure 3.2 shows daily prices (top), daily returns (middle) and TSRV (bottom) of S&P 500 (right) and NASDAQ 100 (left) in the sample period. Table 3.2 reports the summary statistics for the volatility measure of the two series.

The average returns r_t for the two series are close to zero with negative skewness and heavy excess kurtosis. Log-returns standardized by the corresponding realized volatility $r_t/RV_{TS,t}^{1/2}$ show, as observed for instance in Andersen et al. (2001a), an unconditional distribution very close to the Gaussian distribution. Next are reported the summary statistics for the integrated volatilities $RV_{TS,t}^{1/2}$. Both integrated volatilities are positively skewed and extremely leptokurtic. A logarithmic transformation of $RV_{TS,t}^{1/2}$ is sometimes used to approximate a Gaussian distribution, see

Table 3.3: Descriptive statistics for daily log-returns, standardized log-returns and daily realized volatility (4 March 2003 to 29 August 2008)

	Mean	S.d.	Skewness	Kurtosis	Max	Min	JB-test*
			r_t				
S&P500	0.0099	0.7908	0.0325	5.5611	5.2412	-3.1054	0.0010
NASDAQ	-0.0001	1.0733	-0.1814	3.9901	4.1818	-4.2560	0.0010
			$r_t/RV_{TS,t}^{1/2}$				
S&P500	0.0685	1.0345	-0.0615	2.7511	2.7720	-2.9548	0.1272
NASDAQ	0.0378	0.9505	-0.0298	2.8418	3.0077	-2.6032	0.4668
			$RV_{TS,t}^{1/2}$				
S&P500	0.6929	0.3111	2.0987	11.7609	3.3604	0.2104	0.0010
NASDAQ	1.0657	0.3626	1.5188	7.3449	3.4962	0.3706	0.0010
			$\log(RV_{TS,t}^{1/2})$				
S&P500	-0.4488	0.3941	0.4047	3.2464	1.2121	-1.5590	0.0010
NASDAQ	0.0125	0.3152	0.2559	3.3923	1.2517	-0.9928	0.0010

* 0.001 and 0.5 are the smallest and the biggest tabulated p-values for the JB test for small samples in the Matlab function `jbtest.m`

Table 3.4: Descriptive statistics for daily log-returns, standardized log-returns and daily realized volatility (1 September 2008 to 31 October 2008)

	Mean	S.d.	Skewness	Kurtosis	Max	Min	JB-test*
			r_t				
S&P500	-0.4300	3.4621	-0.2333	3.5262	7.4562	-9.2567	0.5000
NASDAQ	-0.6618	3.7929	-0.0955	3.5117	8.5143	-10.0122	0.5000
			$r_t/RV_{TS,t}^{1/2}$				
S&P500	-0.1971	1.1862	0.1223	2.4604	2.7599	-2.2273	0.5000
NASDAQ	-0.2162	1.0602	0.2587	3.0340	2.5039	-2.2750	0.5000
			$RV_{TS,t}^{1/2}$				
S&P500	2.9533	1.5940	1.2198	5.3847	8.8929	0.7690	0.0029
NASDAQ	3.5588	1.6084	0.8974	3.6026	8.6528	1.4519	0.0298
			$\log(RV_{TS,t}^{1/2})$				
S&P500	0.9442	0.5422	-0.1276	2.3951	2.1852	-0.2626	0.5000
NASDAQ	1.1729	0.4455	0.0624	2.0940	2.1579	0.3729	0.3097

* 0.001 and 0.5 are the smallest and the biggest tabulated p-values for the JB test for small samples in the Matlab function `jbtest.m`

Andersen et al. (2003), among others. However, in our case, this transformation is still far away from a Gaussian distribution, due to the positive skewness and the excess of kurtosis still present. As shown in Kim and White (2004), the indexes of skewness and kurtosis are totally unreliable indicators in the presence of extreme events, in the sense that one single extreme observation can jeopardize their meaning. In our sample, such high values for the skewness and kurtosis indicate the presence of abnormally high values for the realized volatility (as a simple look at Figure 3.2 would suggest). This is clearly related to the recent developments in the credit crunch crisis and in particular to the extreme events in the period that goes from 1 September, 2008 until the end of our sample.

To better understand this, we split the sample period in two sub-periods: one of relatively low volatility, and one of very high volatility. The high-volatility sub-sample starts on September 1, 2008. We chose this day as starting point of the VIX index's rise up to its historical maximum level of 80.06 in date 24 October 2008. Figure 3.3 plots the VIX index from 3 March 2008 to 31 October, 2008. Tables 3.3 and 3.4 report the summary statistics for the samples before and after 1 September, 2008. Surprisingly, skewness and excess kurtosis are less present in both the subsamples. In particular, log-returns and standardized log-returns in the extreme-volatility period have a distribution compatible with the Gaussian one. The same holds for the logarithmic transformation of the integrated variance. This means that assuming a single conditional distribution for the realized volatility in the whole sample does not seem the most appropriate choice. We show in the sequel that these different levels in volatility caused by extreme events induce a downward bias in the estimated degrees of freedom of the WAR model.

As previously mentioned, one property of the Wishart distribution is that the volatility of a portfolio follows a gamma distribution. We now show that the gamma distribution is not able to capture extreme events in the volatility process. Using maximum likelihood we fit a gamma distribution to the series of $RV_{TS,t}$. The estimated shape and scale parameter (standard errors are reported in parenthesis) are [0.79 (0.02), 1.19 (0.05)] and [1.15(0.05), 1.29 (0.06)] for the S&P 500 and NASDAQ 100, respectively. Consider the NASDAQ 100. The 99.5% quantile of a Gamma(1.15, 1.29) is equal to 5.7. This means that one expects only 0.5% of the volatility values to be bigger than 5.7. In our sample, however, the percentage of days in which the volatility is bigger than this level is 2.17%, i.e. a gamma distribution fails to detect extreme events in the volatility process.

3.4.2 Analysis with a simulated misspecified WAR

To confirm the findings above, we simulated a sort of miss-specified WAR process that includes extreme observations. We followed the procedure of Section 3.3.2 but assumed that the $x_{t,k}, k = 1, \dots, 7$ in Eq. (3.2.1) come from a fat tail distribution. In particular, using the same matrices

M and Σ as before, we simulated the process $Y_t, t = 1, \ldots, 1000$ as

$$Y_t = \sum_{k=1}^{7} x_{t,k} x'_{t,k} \tag{3.4.4}$$

whit

$$x_{t,k} = M x_{t,k} + \epsilon_{t,k}, \quad k = 1, \ldots, 7.$$

where, differently from (3.2.2), $x_{t,k}$ are not Gaussian vector but are instead assumed to be:

- Sub-Gaussian stable Paretian random vectors[7] with covariance matrix Σ and stability index $\alpha = 2, 1.99, 1.95, 1.9, 1.8$ and 1.7. Recall that the Gaussian distribution is included in the stable family in the special case of $\alpha = 2$. For $\alpha < 2$ the existence of any moment of order higher than 1 is ruled out. In this last case, this distribution possesses fat tails;

- Student's t random vectors with variance/covariance matrix Σ and degrees of freedom $\nu = 3, 5, 10, 20, 30$ and 100. Recall that as the number of degrees of freedom grows, the t distribution approaches the normal distribution. Thus, similarly to the Stable Paretian, low degrees of freedom implies the presence of extreme events in the process.

These two families of distributions have been widely used in the economic literature to capture the excess of kurtosis generally present in financial returns. Stable distributions in finance were introduced by the pioneering works of Mandelbrot (1963) and Fama (1965a). This class of distributions, besides being able to capture extreme events, enjoys many of the properties of the Gaussian, such as closeness under summation, and a number of theoretical results in asset allocation and option pricing are available. See for example Fama (1965a), Fama (1965b), Ortobelli et al. (2002), Ortobelli and Rachev (2005), McCulloch (2003) and the survive by Bradley and Taqqu (2001). For applications to risk modeling see for example Panorska et al. (1995), Mittnik et al. (2000, 2002), Nolan (2003), Haas et al. (2005), Ortobelli and Rachev (2005), Doganoglu and Mittnik (2006), Bonato (2009b). The Student's t distribution (and its skewed version) was also introduced to overcome the incapability of the Gaussian distribution to capture extreme market events. It represents a valid alternative and has a closed-form expression for the density but has the shortcoming of not being closed under summation. This implies a higher degree of difficulty when theoretical results need to be derived. For the application of the Student's t in the financial econometrics literature see for example Bollerslev (1987), Bauwens and Laurent (2005), Giot and Laurent (2003, 2004), Aas and Haff (2006).

The results from the simulations are presented in Table 3.5 (3.6) when Stable Paretian (Student's t) random vectors are used to simulated a misspecified Wishart autoregressive process. Figures 3.4 and 3.5 report the kernel densities of the estimates. Consider Table 3.5 first. When

[7] Sub-Gaussian random vectors are a particular case of symmetric stable Paretian random vector. See Appendix 3.D for the simulation procedure

Table 3.5: The sample mean, the mean of the squared error (MSE) and the mean of the absolute deviation (MAD), sample kurtosis, sample skewness and Jarque-Bera normality test's p-value for the estimators of the degrees of freedom K when a miss-specifien WAR is estimated using stable Paretian random vectors with different stability indexes.

Estim.	Mean	MSE	MAD	Kurtosis	Skewness	J-B test
			$\alpha = 2$			
$\hat{K}_B$	7.0357	0.0924	0.2407	3.0972	0.2027	0.0277
$\hat{K}_G$	7.0523	0.1373	0.2970	3.1324	0.0705	0.4751
			$\alpha = 1.99$			
$\hat{K}_B$	6.3491	1.3521	0.7309	10.3236	-2.3532	0.0010
$\hat{K}_G$	5.1006	8.2276	1.9592	2.8324	-1.0495	0.0010
			$\alpha = 1.95$			
$\hat{K}_B$	4.4684	8.1397	2.5316	3.0401	-0.7808	0.0010
$\hat{K}_G$	1.8518	29.1250	5.1482	2.3653	0.6837	0.0010
			$\alpha = 1.9$			
$\hat{K}_B$	3.1708	15.9216	3.8292	2.4864	-0.3132	0.0010
$\hat{K}_G$	0.7376	39.8928	6.2624	6.1049	1.7156	0.0010
			$\alpha = 1.8$			
$\hat{K}_B$	1.9092	26.4242	5.0908	2.5570	0.1321	0..0010
$\hat{K}_G$	0.2225	46.0064	6.7775	8.8904	2.1718	0.0010
			$\alpha = 1.7$			
$\hat{K}_B$	1.3210	32.5028	5.6790	2.5950	0.2250	0.0010
$\hat{K}_G$	0.1113	47.4742	6.8887	11.1347	2.5175	0.0010

* 0.001 and 0.5 are the smallest and the biggest tabulated p-values for the JB test for small samples in the Matlab function `jbtest.m`

the stability index α is set equal to 2, the simulated process is a Wishart autoregressive process and the two estimators K_B and K_G are unbiased. However, as our previous findings revealed, K_B has a lower variance and a lower MAD. Note that the Jarque-Bera test rejects the null hypothesis of normal distribution for $\hat{K}_B$ at level 5%. Moving down in the table, we see the effect that a smaller α has on the estimation of the degrees of freedom. Already when $\alpha = 1.99$ the two estimators suffer from a downward bias, but still, K_B performs better in terms of MSE and MAD. As α decreases to the lowest level of 1.7 the bias toward zero becomes more severe. For $\alpha = 1.95$ or lower, on average, the estimated degrees of freedom provided by $\hat{K}_G$ are less than 2, the dimension of Y_t minus 1. Thus, according to this estimator, the Wishart distribution does not possess a density function.

Figures 3.13 (3.15) and 3.14 (3.16) plot the estimates and the kernel density of the diagonal elements of M when stable Paretian (Student's t) random vectors are used to simulate the WAR process. We see from these graphics that low values for the stability index α and the degrees of freedom ν have a heavy impact also on the estimation of the matrix M, even though the mode of the distribution for the estimated entries is centered around the true value. This might be a possible explanation for the bias of K_G as it directly depends on $\hat{M}$, but not for the low values of K_B.

As already mentioned, very low values for the estimated degrees of freedom were also found in Chiriac (2006, 2007) when analyzing 8 stocks from the NYSE Trades and Quotes database. In (Chiriac, 2006) this bias is explained as caused by an increase in the sample length of covariance matrices considered and to a high level in the variance of the volatility estimator considered. For the dataset used, since an increase in the series length induces an increase in the variance of the volatility estimators, the performance of the WAR model declines when estimated over samples larger than one month. Chiriac (2007) studies the effect of a cointegrated latent structure of the estimated matrix process. In particular, she relaxes the stationarity assumption on the latent VAR process underlying the WAR process and shows that under (non-stationary) cointegration conditions, the estimated degrees of freedom asymptotically decrease with the cointegration rank and converge in probability to a value smaller than the dimension of the process.

Table 3.6: The sample mean, the mean of the squared error (MSE) and the mean of the absolute deviation (MAS), sample kurtosis, sample skewness and Jarque-Bera normality test's p-value for the estimators of the degrees of freedom K when a miss-specifien WAR is estimated using Student's t random vectors with different degrees of freedom ν.

Estim.	Mean	MSE	MAD	Kurtosis	Skewness	J-B test
			$\nu = 100$			
$\hat{K}_B$	7.0049	0.0902	0.2382	3.1492	0.1383	0.1333
$\hat{K}_G$	7.0128	0.1377	0.2926	3.2050	0.0954	0.2064
			$\nu = 30$			
$\hat{K}_B$	6.9386	0.0894	0.2397	2.9858	0.1764	0.0748
$\hat{K}_G$	6.9243	0.1276	0.2885	2.9536	0.1436	0.1704
			$\nu = 20$			
$\hat{K}_B$	6.8888	0.1055	0.2603	3.0146	0.1416	0.1888
$\hat{K}_G$	6.8452	0.1643	0.3247	2.9008	0.0738	0.5055
			$\nu = 10$			
$\hat{K}_B$	6.6817	0.1891	0.3638	3.0091	0.1282	0.2552
$\hat{K}_G$	6.5570	0.3282	0.4843	3.5949	0.0021	0..0010
			$\nu = 5$			
$\hat{K}_B$	5.8985	1.3208	1.1016	4.0549	-0.2036	0.0010
$\hat{K}_G$	5.1173	4.1101	1.8827	8.3526	-1.7729	0.0010
			$\nu = 3$			
$\hat{K}_B$	3.8053	10.5276	3.1947	5.6762	-1.0941	0.0010
$\hat{K}_G$	1.6653	29.2457	5.3347	2.3101	0.1189	0.0010

* 0.001 and 0.5 are the smallest and the biggest tabulated p-values for the JB test for small samples in the Matlab function `jbtest.m`

In this paper we present a different explanation for the downward bias of the degrees of freedom. We claim this is caused by extreme events in the covolatilities series. In the previous Monte Carlo experiment we relaxed the assumption of gaussianity of the latent VAR processes generating the WAR model. We hypothesized that the low value for the estimated degrees of freedom is not caused by the high variance in the volatility series used in the estimation, but rather by the presence of extreme observations in the process. A criticism to the simulation exercise presented is that the series of matrices we simulated is no longer a WAR process and thus the usual estimators for K will be incorrect. However, in empirical applications, assuming that the process really follows a WAR process is quite an ambitious assumption and it is unlikely to be true. With the previous simulations we intended to deal with processes that offer a more realistic representation of the empirical reality and thus to give a very rough idea of how not accounting for extreme observation in the process affects the estimation of the degrees of freedom.

3.4.3 Analysis with the introduction of outliers

Our next step is to check the impact of outliers on the estimation of the degrees of freedom. We did not relax the Gaussian assumption for the VAR process. Instead, we introduced outliers in the series in order to induce extreme observations compatible with the distribution of assets realized volatility.

One of the earliest papers on the detection and testing of outliers in stationary time series is that by Fox (1972). He introduced and defined outliers that are additive (AO) or innovative (IO). The paper presented cases in which problems were caused by having either all AO or all IO types presented in the time series. Chernick et al. (1982) demonstrated a way in which the presence of outliers at isolated time points can influence autocorrelations at different lags and proposed a visual procedure for detecting them. Muirhead (1986) discussed a method of distinguishing outliers types in an autoregressive (AR) process. More recently Chen and Liu (1993) investigated outlier issues in time series forecasting using a detection and adjustment approach.

Given a simple stationary AR(1) process of the form

$$x_t = \rho x_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma 2), t = 1, \ldots, T \tag{3.4.5}$$

and a random time $\tau \in (1, T)$, an AO with size x_{out} is added, once the process x_t is simulated, by setting $x_\tau = x_{out}$. In Cai and Davies (2003), for example, they set $\rho = 0.5, \sigma = 0.2$ and $x_{out} = 0.7$, i.e. the outlier lies outside three standard deviation from the mean of the process. The same logic applies when introducing the innovation outlier. However, in this case, the other $x_t, t = \tau + 1, \ldots, T$ are obtained from Equation (3.4.5) by using the outlier value to produce the remaining data. An alternative way to introduce an innovative outlier (see for example Battaglia, 2005) is to write

$$x_\tau = x_{\tau-1} + \eta_\tau \tag{3.4.6}$$

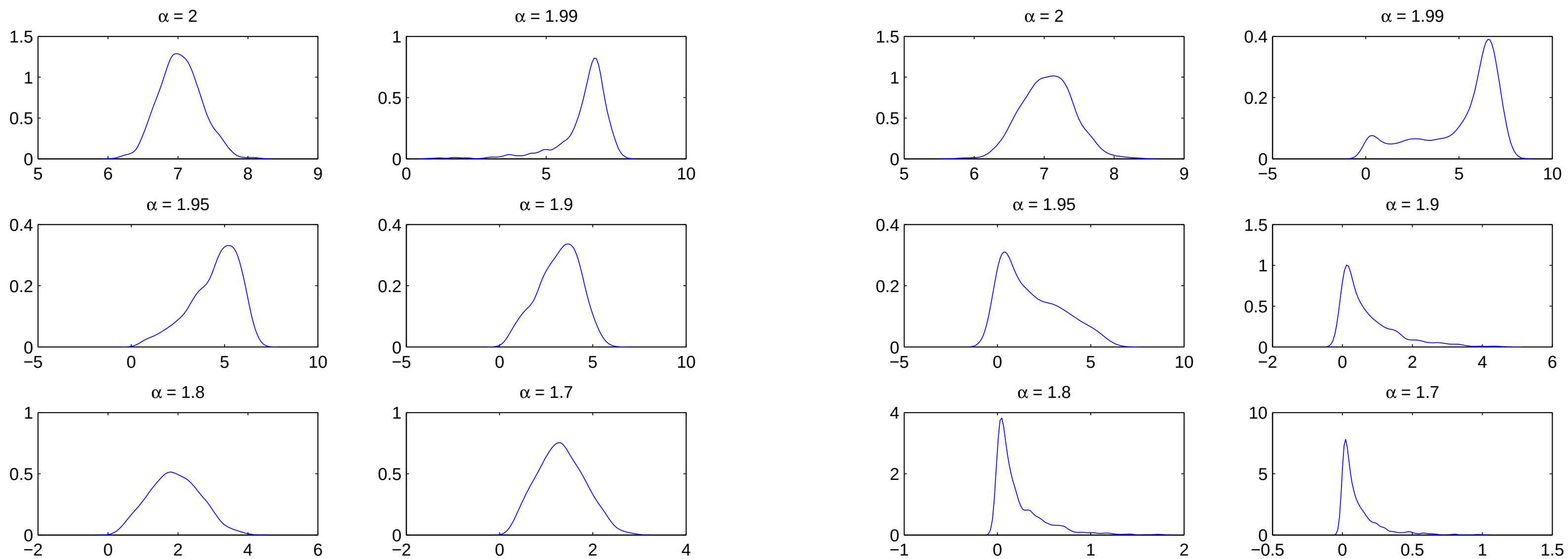


Figure 3.4: Kernel densities of $\hat{K}_B$ (left group) and $\hat{K}_G$ (right group) when Sub-Gaussian random vector with different stability indexes α are used to simulated the WAR(1).

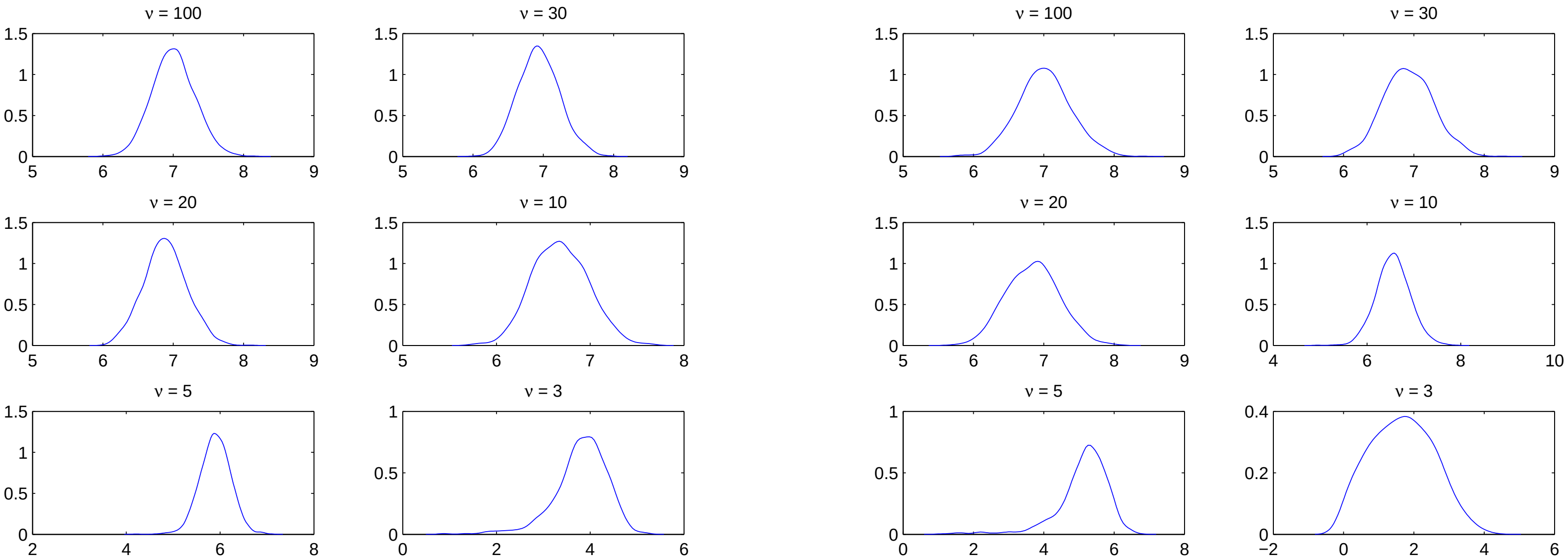


Figure 3.5: Kernel densities of $\hat{K}_B$ (left group) and $\hat{K}_G$ (right group) when Student's t random vector with different degrees of freedom ν are used to simulated the WAR(1).

where $\eta_\tau = \epsilon_\tau + \omega$ and ω is the noise that induces the innovative outlier. In this simulation study we added one, two or three (additive or innovative) outliers.

To generate the *additive* outliers we followed Kim and White (2004) and transferred their procedure to the matrix case. Outliers are constructed to occur at a random time $\tau_i \in (0,1), i = 1,2,3$[8]. From the S&P 500 and NASDAQ 100 tick-by-tick futures prices previously introduced, we obtained the series of equally-spaced prices at a five-minute frequency using the previous-tick interpolation method. The series of realized covariance matrices were obtained using the classical estimator presented in Andersen et al. (2003):

$$Y_t = \sum_{i=1}^{I} r_{t-1+ih,h} r'_{t-1+ih,h} \tag{3.4.7}$$

where p_{t-1+ih} denotes the $(n \times 1)$ vector of log-close transaction prices, $r_{t-1+ih,h} \equiv p_{t-1+ih} - p_{t-1+(i-1)/h}$ denotes the $(n \times 1)$ vector of returns for the i-th intraday period on day t, for $i = 1, \ldots, I$. $n = 2$ is the number of stocks. I is the number of intraday intervals, each of length $h \equiv 1/I$. In our case, with a frequency of five minutes, $I = 81$. In contrast to de Pooter et al. (2006) we did not consider overnight returns. We indicate with Y_t the realized covariance matrix at time t in order to to be coherent with our previous notation and because the use of Σ would probably create confusion as Σ denotes the covariance matrix of the Gaussian vector underlying the WAR(1) model.

The detection and simulation of outliers in a series of WAR matrices is, to our knowledge, still an untouched research field. As the goal of multivariate statistics in Finance is to model a group of asset composing a portfolio, we rely on an equally-weighted portfolio as auxiliary series to find the size of the outliers in the series of covariance matrices.

To find the "outlier matrices" in our sample we first computed the volatility of the equally weighted portfolio of the two series of realized volatilities i.e.

$$RV_t(\omega) = \omega' Y_t \omega, \quad t = 1, \ldots, 1344 \tag{3.4.8}$$

where ω is the vector of equal weights. From the three largest observations in $RV_t(\omega)$, with values 34.86, 35.66 and 58.42 and location τ_1, τ_2 and τ_3 (8, 16 and 10 October 2008), we recovered the corresponding three matrices $Y_{\tau_1}, Y_{\tau_2}, Y_{\tau_3}$ and considered them as the "outlier" matrices. Again, from $RV_t(\omega)$, we calculated the 25th percentile of its sampling distribution and extracted the corresponding matrix in Y_t, denoted by $Y_t^{.25}$. The size of the outlier matrix relative the 25th

[8]Differently from Kim and White (2004), in which the location of the outlier in the simulated series correspond the the location of the biggest absolute observation in the series of daily returns on the S&P 500 index, we used a random timing. This is because in our samples, extreme observations are clustered at the end of the period and this is not always the general case.

percentile is computed as:

$$\begin{aligned} Y_{m_1} &= Y_{\tau_1}./Y_t^{.25} \\ Y_{m_2} &= Y_{\tau_2}./Y_t^{.25} \\ Y_{m_3} &= Y_{\tau_3}./Y_t^{.25} \end{aligned}$$

where ./ denotes the elementwise ratio. Then, we generate random WAR processes $\{Y_t\}_{t=1,\dots,N}$ and calculate the 25th percentile $F^{-1}(0.25)$ of the corresponding portfolio volatility $RV_t(\omega)$ to get $Y_t^{.25}$. The i-th additive outlier is $Y_{m_i}.*Y_t^{0.25}$, where $.*$ denotes the elementwise product.

To introduce *innovative* outliers we added noise in the Gaussian VAR processes that generate the WAR systems. Following the lines of Battaglia (2005) but in a multivariate multiplicative fashion, at a random time τ we simulate the VAR(1) processes as in Equation (3.2.2):

$$x_{k,\tau} = M\tilde{x} + \epsilon_{k,t}, \;\; k = 1, \dots, k \tag{3.4.9}$$

with $\tilde{x} \sim N(0, \delta\Sigma)$ and the noise component δ assumes values 9,6 and 3 for the one, two or three outliers we add.

We follow the previous Monte Carlo experiment and simulate 1,000 WAR(1) processes consisting of 1,000 observations each. The values for M, Σ and K were calibrated on the estimates obtained from the sequence of covariance matrices for the S&P 500 and NASDAQ 100. To simplify the estimation procedure in the simulation we assumed M to be diagonal. The period considered for the estimation goes up to 29 August 2008. i.e. the relatively low volatility period. The values for M and Σ are

$$\widehat{M} = \begin{pmatrix} 0.79 & 0 \\ 0 & 0.77 \end{pmatrix} \text{ and } \widehat{\Sigma} = \begin{pmatrix} 0.08 & 0.09 \\ 0.09 & 0.2 \end{pmatrix}.$$

The estimated degrees of freedom are $\hat{K}_B = 4.25$ and $\hat{K}_G = 2.71$ and we chose $K = 3$.

Figure 3.7 plots the series $Y_t(1,1)$, i.e. the simulated realized volatility for the first asset, with 1,2 or three additive (left panels) or innovative (right panels) outliers. Additive outliers are bigger in magnitude than innovative outliers and thus we expect them to have a greater impact on the degrees of freedom estimates.

Table 3.7 reports mean, MSE, MAD, skewness, kurtosis and JB test for the estimated degrees of freedom K using the estimators K_B and K_G when one, two or three additive or innovative outliers are added to the simulated process. Figure 3.6 shows the kernel densities of the two estimators when outliers are introduced. In front of a theoretical value of 3 degrees of freedom, we clearly see that $\hat{K}_B$ is less affected by the presence of extreme observations than $\hat{K}_G$. However, both estimators suffer from bias toward zero induced by the presence of outliers. This results are confirmed in Table 3.7. When innovative outliers are introduced, the mean of $\hat{K}_B$ ranges

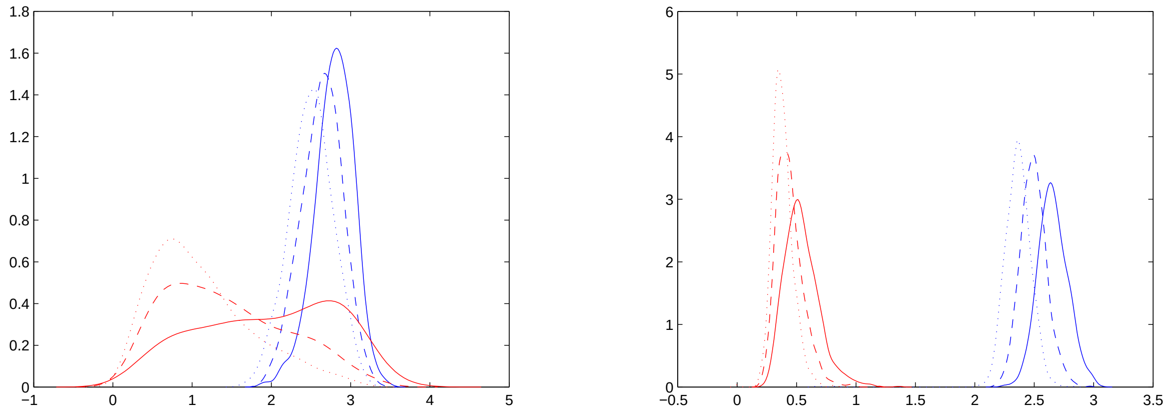


Figure 3.6: Kernel densities of $\hat{K}_B$ (blue lines) and $\hat{K}_G$ (red lines) when one, two or three (solid, dashed and dotted line) innovative (left) or additive (right) outliers are added to the simulated WAR process

from 2.79 to 2.50 for one and three outliers, respectively. The downward bias is more severe for $\hat{K}_G$ with average values from 1.96 to 1.13. Mean square error and mean absolute error indicates that $\hat{K}_B$ provides better results when compared with $\hat{K}_G$. Note also that at level 1%, we cannot reject the hypothesis of normal distribution for $\hat{K}_B$ and the distributions of $\hat{K}_G$ are far from being normal and are extremely skewed. The second part of the table shows the same statistics when additive type outliers are added to the simulated process. As previously said, the impact of additive outliers is expected to be heavier compared to innovative outliers. This is indeed confirmed by the average values of the two estimators. However, while for $\hat{K}_B$ this implies a shift to 2.65, 2.49 and 2.33 for one, two or three outliers, the average values of $\hat{K}_G$ collapse to mean values much smaller, i.e. 0.54, 0.44 and 0.38. Again, MSE and MAD for $\hat{K}_B$ are slightly worse but still in line with the innovative outliers case. On the contrary, when $\hat{K}_G$ is used to estimated the degrees of freedom, MSE and MAD present higher value if compared with the previous case. This is somehow expected, given the larger downward bias of this estimator.

Again, Figure 3.17 and 3.18 show the estimated diagonal entries of M and their kernel densities. Additive outlier have a severe impact and the estimates suffer from a downward bias. Innovative outliers have a mild impact in term of downward bias and induce a light left skewness on the distribution of the estimates.

This simulation exercise helped us to confirm the previous findings. In particular we showed that, first, the presence of extreme events in the process, whether they are implied by the data generating process (Stable and Student's t for instance) or they are simply outliers, causes a bias toward zero of the estimated degrees of freedom. Second, independently of the type of outliers, the estimator that relies on the gamma distribution, $\hat{K}_B$, is preferable according to the MSE and MAD criteria.

Table 3.7: The sample mean, the mean of the squared error (MSE) and the mean of the absolute deviation (MAS), sample skewness, sample kurtosis and Jarque-Bera normality test's p-value for the estimators of the degrees of freedom K when additive or innovative outliers are introduced in the process.

Estim.	Mean	MSE	MAD	Skewness	Kurtosis	J-B test
Innovative Outlier						
1 Outlier						
$\hat{K}_B$	2.7990	0.0999	0.2462	-0.4817	3.5793	0.0010
$\hat{K}_G$	1.9899	1.7810	1.0486	-0.3307	1.9302	0.0010
2 Outliers						
$\hat{K}_B$	2.6507	0.1918	0.3662	-0.2526	3.1305	0.0068
$\hat{K}_G$	1.4980	2.8228	1.5143	0.3775	2.3320	0.0010
3 Outliers						
$\hat{K}_B$	2.4772	0.3466	0.5271	-0.0117	2.7460	0.2436
$\hat{K}_G$	1.0881	4.0535	1.9132	0.9187	3.1745	0.0010
Additive Outlier						
1 Outlier						
$\hat{K}_B$	2.6573	0.1330	0.3428	0.1636	3.0893	0.0849
$\hat{K}_G$	0.5446	6.0507	2.4554	0.9784	4.8763	0.0010
2 Outliers						
$\hat{K}_B$	2.4945	0.2675	0.5055	0.2993	3.4140	0.0010
$\hat{K}_G$	0.4424	6.5549	2.5576	1.2035	6.1384	0.0010
3 Outliers						
$\hat{K}_B$	2.3622	0.4226	0.6378	-4.1484	57.0325	0.0010
$\hat{K}_G$	0.3822	6.8611	2.6176	0.7904	5.6411	0.0010

* 0.001 and 0.5 are the smallest and the biggest tabulated p-values for the JB test for small samples in the Matlab function `jbtest.m`

3.5 Modeling extreme volatility risk

The previous sections shed some light on the effect that extreme observations in the variance-covariance process have on the estimated degrees of freedom. In particular we have seen that the introduction of shocks in the matrix Σ (innovative outliers) induces a downward bias in the estimated degrees of freedom.

In this section we first use the method of moments estimator to determine the role of Σ in the estimation of degrees of freedom K. Then, following the lines of Bertholon et al. (2009), we introduce a mixture of Wishart to model extreme events in the volatility process. We show that this type of model is able to explain the downward bias for the estimated degrees of freedom.

Earlier in this paper we showed that a more efficient estimator K can be obtained by simply fitting, via maximum likelihood, a gamma distribution to the volatility of any portfolio. To see the effect of shocks in the volatility of volatility (represented by the matrix Σ), we adopt a

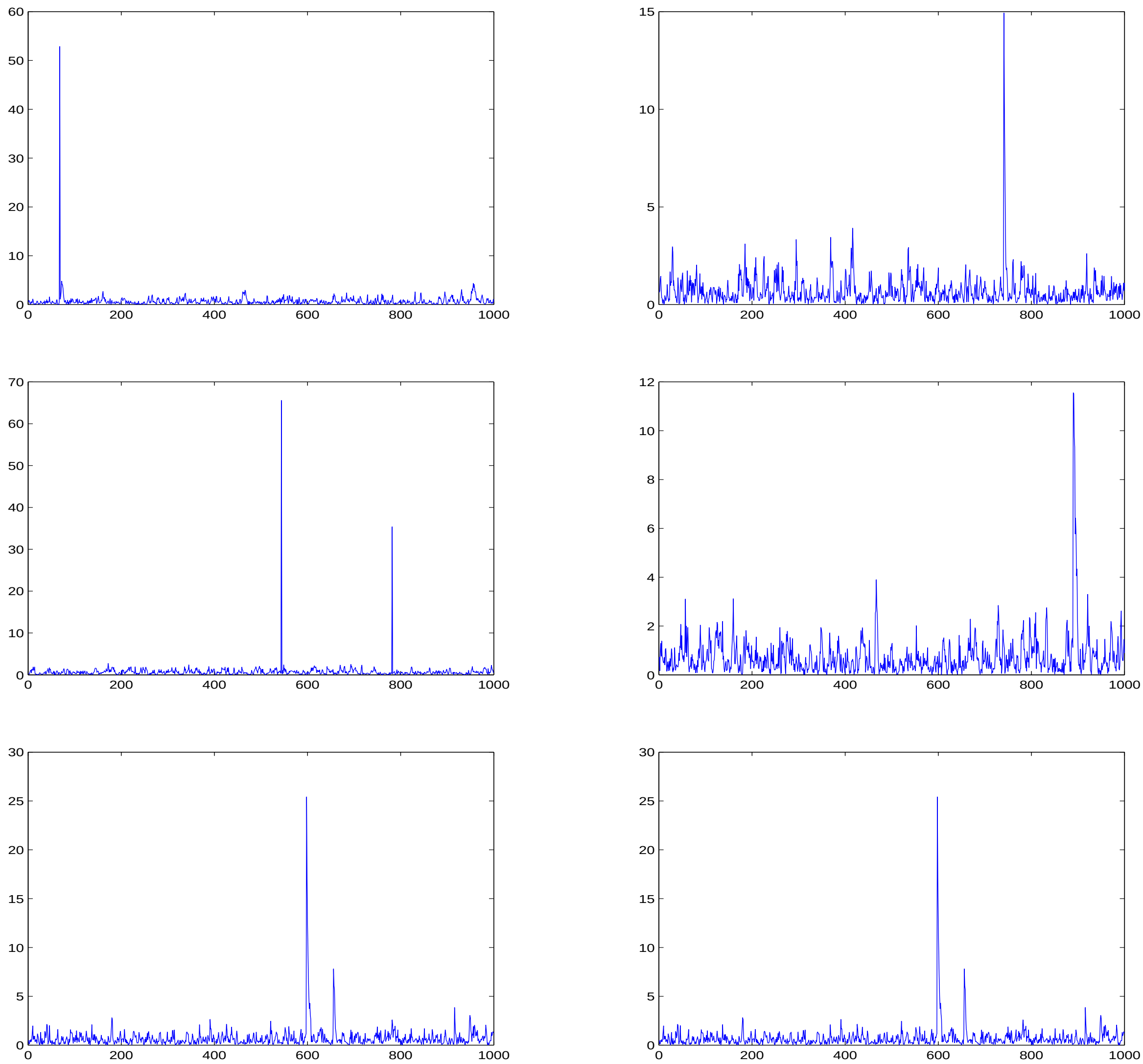


Figure 3.7: Simulated realized volatility of the first asset (i.e. Yt(1,1)) with 1, 2 or 3 additive (left panels) or innovative (right panels) outliers.

third estimator to compute the parameter of interest: the method of moments estimator. This estimator, although less efficient than the MLE, has the advantage of possessing a closed form expression for the parameters of the gamma distribution.

Recall that for any matrix $Y \in \mathbb{R}^{n\times n}$ with centered Wishart distribution $W(K, \Sigma)$ and any vector $\alpha \in \mathbb{R}^n$, we have that $Z = \alpha' Y \alpha \sim Ga(K/2, 2\alpha'\Sigma\alpha)$. Then:

$$E[Z] = K(\alpha'\Sigma\alpha) \tag{3.5.1}$$

$$V[Z] = 2K(\alpha'\Sigma\alpha)^2 \tag{3.5.2}$$

The method of moments estimator, say $\hat{K}_{mm}$, is simply:

$$\hat{K}_{mm} = \frac{2\hat{E}[Z]^2}{\hat{V}[Z]}$$

Thus, shocks in the variance of the variance of the portfolio induce a downward bias in the estimated degrees of freedom. Note also that $V[Z] = 2(\alpha'\Sigma\alpha)^2$ so any shock in $V[Z]$ is directly imputable to shocks in Σ.

To account for perturbations of a WAR process we introduce a mixture model similar to the one presented, using normal distributions, in Bertholon et al. (2009). We assume that the density of the variable Y (Y_t for the conditional distribution), representing the covariance matrix of n assets, takes the form:

$$f(Y) = pW\left[K, \frac{\Sigma}{2p}\right] + (1-p)W\left[K, \frac{\Sigma}{2(1-p)}\right] \tag{3.5.3}$$

$$f(Y_t|\mathbb{I}_{t-1}) = pW\left[K, M, \frac{\Sigma}{2p}\right] + (1-p)W\left[K, M, \frac{\Sigma}{2p}\right]. \tag{3.5.4}$$

where $p \in \,]0,1[$, $W[K,\Sigma]$ and $W[K,M,\Sigma]$ represent the density of a centered and of an autoregressive Wishart distribution, respectively.

The intuition behind this model is that each realization Y (or Y_t) comes from a mixture of two Wishart distributions, one of which having a small weight associated with a strong variance represented by the matrix Σ. As we recover the degrees of freedom fitting a Gamma distribution to the volatility of a portfolio, rather than studying directly Y, we are interested in the characteristics of the distribution of $\alpha'Y\alpha$ (or $\alpha'Y_t\alpha$). In particular, we want to check the behavior of $V[\alpha'Y\alpha]$ depending on the value of p as, recall, this affecta the estimation of the degrees of freedom.

Proposition 2. *The main characteristics of this distribution are:*

Unconditional case:

(u1) $E[\alpha'Y\alpha] = K\alpha'\Sigma\alpha$

(u2) $V[\alpha'Y\alpha] = K(\alpha'\Sigma\alpha)^2\left(\frac{K+2}{4p(1-p)} - K\right)$

(u3) $Y \sim W(K,\Sigma) \quad if \quad p = 1/2$

(u4) $Y \overset{D}{\nrightarrow} W(K,\Sigma/2) \quad as \quad p \to 0.$

Conditional case:

(c1) $E[\alpha'Y_t\alpha|\mathbb{I}_{t-1}] = \alpha'MY_tM'\alpha + K\alpha'\Sigma\alpha$

(c2) $V[\alpha'Y_t\alpha|\mathbb{I}_{t-1}] = 4\alpha'MY_tM'\alpha'\Sigma\alpha + (\alpha'\Sigma\alpha)^2\frac{K}{2}\left(\frac{1}{p(1-p)} + \frac{K}{2p(1-p)} - K\right)$

(c3) $Y_t|\mathbb{I}_{t-1} \sim W(M,K,\Sigma) \quad if \quad p = 1/2$

(c4) $Y_t|\mathbb{I}_{t-1} \overset{D}{\nrightarrow} WAR(M, K, \Sigma/2)$ *as* $p \to 0$.

[Proof: see Appendix 3.C]

The intuition behind this mixture model is slightly different from the one in Bertholon et al. (2009). In our case, any deviation from $p = 1/2$ induces a shock in the variance of the process but does not affect the mean value. In the extreme case where $p \to 0$ the variance of the process explodes and does not exist finite. Thus in our formulation, any time the probability p moves from the value of 1/2, the model results contaminated by more and more extreme shocks in the variance process.

To see the effect that the value of p has on the estimation of the degrees of freedom, we first use the methods of moments estimator of K, K_{mm}. We plot the values of the degrees of freedom as function of p under the wrong assumption that the process Y follows a $W(K, \Sigma)$, i.e. when $p \neq 1/2$. The method of moments estimator under the wrong assumption of no mixture reads:

$$\begin{aligned}
\hat{K}_{mm} &= 2\frac{\hat{E}[\alpha' Y \alpha]^2}{\hat{V}[\alpha' Y \alpha]} \\
&= \frac{2(K\alpha'\Sigma\alpha)^2}{K(\alpha'\Sigma\alpha)^2\left(\frac{K+2}{4p(1-p)} - K\right)} \\
&= \frac{2K}{\frac{K+2}{4p(1-p)} - K} \to 0 \quad \text{as} \quad p \to 0 \quad \text{or} \quad p \to 1.
\end{aligned}$$

Figure 3.8 (left) plots for $p \in \ ]0, 1[$ the different estimates of K, whose true value is set to be 7, obtained using the method of moments estimator under the incorrect assumption that $p = 1/2$. As p moves from 1/2 toward either 0 or 1, the estimated degrees of freedom are, as expected, biased toward zero.

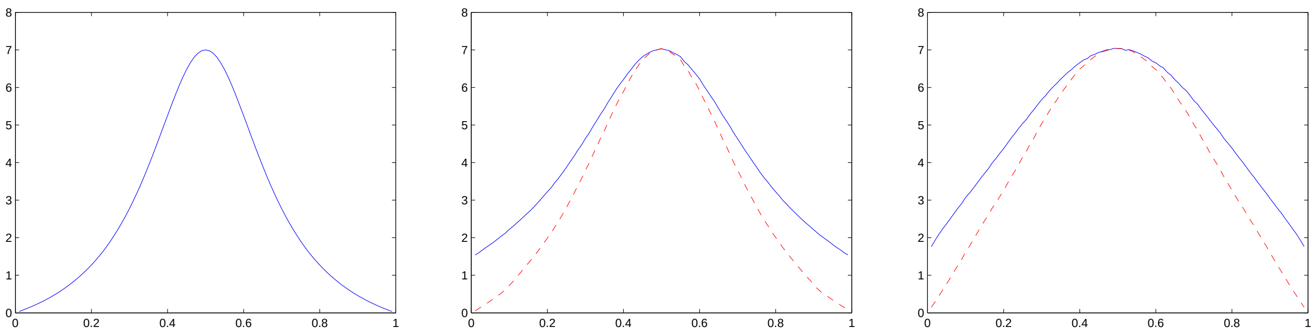


Figure 3.8: For different values of $p \in \ ]0, 1[$ and $K = 7$, corresponding values of K_{mm} when K is known (left); $\hat{K}_B$ (blue) and $\hat{K}_G$ (red dashed) when a $W(K, \Sigma)$ (center) and a $WAR(K, M, \Sigma)$ (right) are simulated.

In this last example we used the method of moments estimator. This estimator has the advantage of having a closed form expression for the degrees of freedom K_{mm} but has the disadvantage of being less efficient than the ML estimtor, which is the one we use in practice. The ML estimator, however, does not have a closed form expression for $\hat{K}_B$ so we rely on simulations to see the effect of $p \neq 1/2$ when a mixture of Wishart is simulated. For $p \in \,]0,1[$ we simulated 1,000 simple paths coming from (3.5.3) and (3.5.4). Each simulated series has sample size 1,000, $K = 7$, $M = diag([0.5\ 0.4\ 0.2\ 0.8])$ and $\Sigma = diag([1\ 1\ 1\ 1])$. For different values for p we estimated the degrees of freedom. Figure 3.8 (center and right) plots the mean values of the estimated $\hat{K}_B$ as p varies (blue solid line). The red dashed line represents the estimated degrees of freedom using the estimator K_G. The pattern is very similar to the one of the previous example and confirms that any deviation of p from the value 1/2 induces a bias toward zero caused by bigger and bigger shocks in the volatility of the process. Note also that, again, K_G and K_B are both unbiased when the process is a true Wishart (or WAR) but $\hat{K}_G$ collapse to 0 quicker as the process becomes more and more contaminated by extreme events.

3.6 Empirical Application

The last section of the paper is dedicated to an intensive empirical analysis of the estimation of the degrees of freedom for a WAR model applied to the sequence of realized variance-covariance matrices of the pair S&P 500-NASDAQ 100 futures.

As shown in Gourieroux et al. (2009) and, for a restrict parametrization, in Bonato et al. (2008), the WAR process is a suitable tool to model and forecast the realized covariance matrix of a group of assets. It is easy to implement, it guarantees the forecasted matrix to be positive definite and the coefficients of the model are directly interpretable. Our goal is to investigate the behavior of the estimated degrees of freedom when different estimators and different sampling frequencies are adopted to compute the series of realized variance-covariance matrices.

To estimate the variance-covariance matrix, we used three estimators among the ones proposed in literature: the classical estimator as in Andersen et al. (2003) previously introduced, two times scales estimator and a kernel-type estimator, both proposed in de Pooter et al. (2006). The sampling frequencies adopted are: 15 seconds, 1, 2, 3, 5, 10, 15, 30, 65 and 130 minutes. The CME trading floor is open from 8.30 to 15.30 and we discharged the first 15 minutes of observation in order to eliminate the overnight effect and this left us with a 6,5 hours trading day.

The first estimator of the realized covariance matrix, the standard estimator, $V_{t,h}$ reads:

$$V_{t,h} = \sum_{i=1}^{I} r_{t-1+ih,h} r'_{t-1+ih,h}, \tag{3.6.1}$$

$r_{t-1+ih,h}$ denotes the i-th intraday returns computed as explained in Section 3.4.3. The second estimator is a multivariate generalization of the two time scales realize volatility estimator proposed by Zhang et al. (2005) adopted in Section 3.4. The two time scale estimator $V_{t,h}^{TTS}$ as defined in de Pooter et al. (2006) is obtained as

$$V_{t,h}^{TTS} = \frac{I_{max}}{I_{max} - I} \left(V_{t,h}^{SubS} - \frac{I}{I_{max}} V_{t,h}^{max} \right), \tag{3.6.2}$$

where $V_{t,h}^{SubS}$ is the subsampling estimator when using I returns and $V_{t,h}^{max}$ is the realized covariance matrix based on the highest possible sampling frequency with I_{max} intraday return observations. As for the univariate case, our highest frequency is 15 seconds.

The last estimator we implement is obtained by adding lead and lagged covariances to the contemporaneous cross-product of returns. This might help to reduce the downward bias in the realized covariances and reduces the upward bias in the realized variance due to the negative autocorrelations in high-frequency returns, see Hansen and Lunde (2005, 2006). Following de Pooter et al. (2006), let $\Gamma_{t,h,l}$ denote the l-th cross-covariance matrix of intraday h-period returns, that is,

$$\Gamma_{t,h,l} = \sum_{i=l+1}^{I} r_{t-1+ih,h} r'_{t-1+(i-l)h,h}. \tag{3.6.3}$$

The realized covariance matrix with lead and lags is then obtained as

$$V_{t,h}^{LL} = V_{t,h} + \sum_{l=1}^{q} d_l (\Gamma_{t,h,l} + \Gamma'_{t,h,l}), \tag{3.6.4}$$

where $V_{t,h}$ is as in Equation (3.6.1) and the weights for the leads and lags are taken to be $d_l = 1 - l/(q+1)$, i.e. Bartlett-kernel weights.

For each series of estimated realized covariances matrices computed at different sampling frequencies, we estimated the degrees of freedom using the two estimators. Results, along with the mean and variance of the realized variances and realized covariances, are reported in Table 3.8. For the three estimators of the covariance matrix, the standard, the two time-scales and the kernel, we see that the average realized variance increases with the sampling frequency and the the average realized covariances decreases with the sampling frequency. These are common pattern widely documented in literature. The variance of the realized variance has a less clear pattern. For the standard and the kernel estimator, on average, the sample variance decreases with the sampling frequency and stabilizes at the 10-15 minutes frequency. Then it increases when data are sampled more sparsely. In general, one expects the variance of the realized volatility become smaller for higher frequencies simply because more data point are used in the estimation. The fact that the pattern is the opposite for frequencies from 15 seconds to 5 minutes might be due to the effect of the market micro-structure noise. In the case of the two time-scales estimator, we have a decrease in mean and variance of the realized volatility as the sampling frequency decreases. For the kernel and the standard estimator, the mean values of the realized volatility, range between 1.6 and 1.2. With frequencies higher than 30 minutes, the average variance computed used the two time-scales estimator is 0.9 and 0.6. This very low values for the average realized variance, along with the low variance of this estimator, require further attention which is beyond the goal of this paper[9].

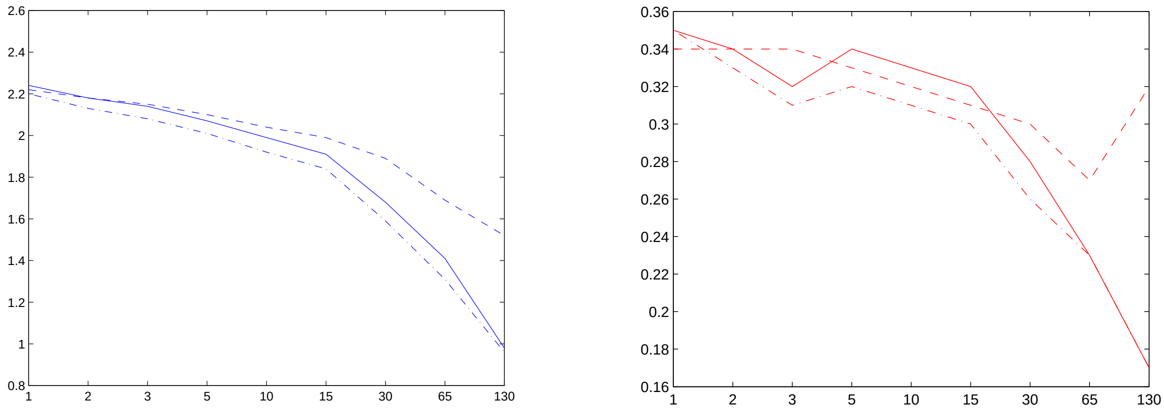


Figure 3.9: Estimated degrees of freedom using K_B (left) and K_G (right) as function of the sampling frequency and of the estimator for realized covariances: standard (solid line), two time-scale (dashed line) and kernel (dash-dot line) .

The first two columns of Table 3.8 report the real object of interest: the estimated degrees of freedom. Figure 3.9 plots the estimated degrees of freedom as function of the sampling frequencies when for different estimators of the covariance matrices. As expected, the estimated degrees of freedom using the gamma distribution are always larger than the values returned by the other estimator. In our simulation experiment we found that K_B is much less affected than K_G by extreme events and this finding is confirmed here. In particular, for all the different sampling frequencies, K_B is on average (excluding two cases) bigger than $n-1=1$, i.e. the Wishart process is always defined. For sampling frequencies higher than 10 minutes, it is always bigger than $n=2$. i.e. the Wishart process is non-degenerate.

All the K_G estimates are, on average, lower than 1, i.e. according to this estimator, the Wishart process does not possess a density. A common, interesting, feature of both estimators, is that, first, the degrees of freedom increase with the sampling frequency. Second, the degrees of freedom are not influenced by the variance of the realized volatility: in front of a 'U' shape of the variance of the realized volatility as function of the sampling frequency, the estimated degrees preserve a decreasing pattern.

3.6.1 WAR with a rolling window

Fitting a single model to capture the conditional distribution of a sample over a long time period might be a quite restrictive choice as different regimes can be present in the variance covariance

[9]The same analysis has been performed on an alternative dataset, kindly provided by Angelo Ranaldo, consisting of the same series of S&P 500 and NASDAQ 100 from 1997 to 2003 at a 5-minute frequency . The figures did not present notable differences.

Table 3.8: The table reports the estimated degrees of freedom using K_B and K_G and mean and variances of the realized (co)-variances for different sampling frequencies for S&P 500 and NASDAQ 100 indexes from March 3, 2003 to October 31, 2008. For the realized variance, the mean reflects the average taken over the two indexes and all over the 1344 trading days. The variance is the average of the two sample variances of the realized variances. In the Panel A the 'standard' realized covariance matrix $V_{t,h}$ given in (3.6.1) is used. The two time estimator $V_{t,h}^{TTS}$ as in (3.6.2) is used in Panel B, while Panel C shows results for the lead-lag corrected estimator $V_{t,h}^{LL}$ given in (3.6.4), with Bartlett-kernel weights $d_l = 1 - l/(q+1)$ and $q = 1$.

		Degrees of freedom		Realized Variance		Realized Covariance	
		K_B	K_G	mean	variance	mean	variance
Panel A: Standard							
15	seconds	2.22	0.4	1.655	16.328	0.324	0.412
1	minutes	2.24	0.35	1.504	13.644	0.648	2.584
2	minutes	2.18	0.34	1.443	12.273	0.787	4.181
3	minutes	2.14	0.32	1.408	12.527	0.852	5.217
5	minutes	2.07	0.34	1.339	10.755	0.905	5.371
10	minutes	1.99	0.33	1.262	9.439	0.966	6.299
15	minutes	1.91	0.32	1.231	9.378	0.980	6.724
30	minutes	1.68	0.28	1.223	10.295	1.023	8.537
65	minutes	1.41	0.23	1.177	11.217	1.018	9.903
130	minutes	0.98	0.17	1.241	17.817	1.092	16.089
Panel B: Two time-scale							
15	seconds						
1	minutes	2.22	0.34	1.474	13.301	0.741	3.362
2	minutes	2.18	0.34	1.410	11.925	0.845	4.697
3	minutes	2.15	0.34	1.372	11.209	0.893	5.278
5	minutes	2.1	0.33	1.316	10.466	0.937	5.943
10	minutes	2.04	0.32	1.229	9.278	0.959	6.397
15	minutes	1.99	0.31	1.181	8.582	0.959	6.543
30	minutes	1.89	0.3	1.071	7.283	0.909	6.208
65	minutes	1.69	0.27	0.903	5.743	0.789	5.225
130	minutes	1.52	0.32	0.643	2.441	0.568	2.226
Panel C: Kernel							
15	seconds	2.25	0.39	1.595	14.949	0.479	1.105
1	minutes	2.2	0.35	1.440	12.121	0.788	4.065
2	minutes	2.13	0.33	1.372	11.309	0.893	5.439
3	minutes	2.08	0.31	1.329	11.049	0.933	6.256
5	minutes	2.01	0.32	1.269	9.841	0.963	6.444
10	minutes	1.92	0.31	1.218	9.351	0.995	7.267
15	minutes	1.84	0.3	1.203	9.277	1.003	7.452
30	minutes	1.59	0.26	1.224	10.914	1.053	9.637
65	minutes	1.31	0.23	1.193	12.276	1.042	10.928
130	minutes	0.96	0.17	1.229	17.336	1.081	15.427

process and parameters may be time varying. A simple solution to this problem is to use a rolling window. This is done for example by finance practitioners for risk management purposes. To check the variation over time of the degrees of freedom we estimated the WAR model using a 21 and 62 (trading) days window that covers the entire sample in analysis. Figures 3.10 and 3.11 show the estimated degrees of freedom using K_B (blue line) and K_G (red line). The sequence of realized covariance matrices was constructed using the standard estimator with sampling frequency 15 seconds (top left), 5 minutes (top right), 30 minutes (bottom left) and 65 minutes (bottom right). The four plots confirm that a lower sampling frequency induces lower estimates for the degrees of freedom. In general, the two estimators look similar and they virtually coincide. K_B, at a 15 seconds and 5 minutes sampling frequency is always bigger than 2 (black line in the plots) for every windows whereas this is not true for K_G. From this rolling windows experiment we conclude that, first, degrees of freedom are likely to vary over time. Second, the values of the degrees of freedom seem to be a function of the sampling frequency at which the series of realized covariances is constructed and not of the length of the estimation window. Note also how, to some extent, the values of the estimated degrees of freedom drop when high peaks in the volatility process are present. This confirms our previous findings.

3.7 The WAR under cointegration assumptions

As stated in Chiriac (2007), one possible explanation for the low values of the estimated degrees of freedom is that the Wishart process is not stationary and a non-stationary specification of the WAR model il considered. More precisely, the processes $x_{k,t}$ of dimension $n \times 1$, with $k = 1, \ldots, K$ are assumed to be cointegrated with cointegration rank $r, r < n$.

For each cointegrated process $x_{k,t}$, its vector error correction (VEC) form is

$$\Delta x_{k,t} = -Hx_{k,t-1} + \epsilon_{k,t}, \quad \epsilon_{k,t} \sim N(0, \Sigma), \tag{3.7.1}$$

where H is $n \times n$ and has rank $rank(H) = r$, i.e. is not of full rank, and the autoregressive matrix M becomes $M = \mathbf{I}_n - H$. Given that H has rank r, it can be written as the product of the two matrices B and Γ, both of dimension $n \times r$ and full rank r: $H = B\Gamma'$. The process $x_{k,t}$ has then the autoregressive representation

$$x_{k,t} = (\mathbf{I}_n - H)x_{k,t-1} + \epsilon_{k,t}. \tag{3.7.2}$$

This process is defined in Chiriac (2007) a ‘nonstationary Wishart autoregressive process of order 1’, NoWAR(1). For this model, an estimator of the degrees of freedom K, is derived and reads (see Appendix A.2 in Chiriac, 2007 for derivation):

$$\hat{K} = \frac{\sum_{k=1}^{K} \sum_{i=1}^{r} \sum_{t=1}^{T} z_{k,i,t}^2}{T^2 \cdot \mathrm{Tr}(C'\hat{\Sigma}C)}; \tag{3.7.3}$$

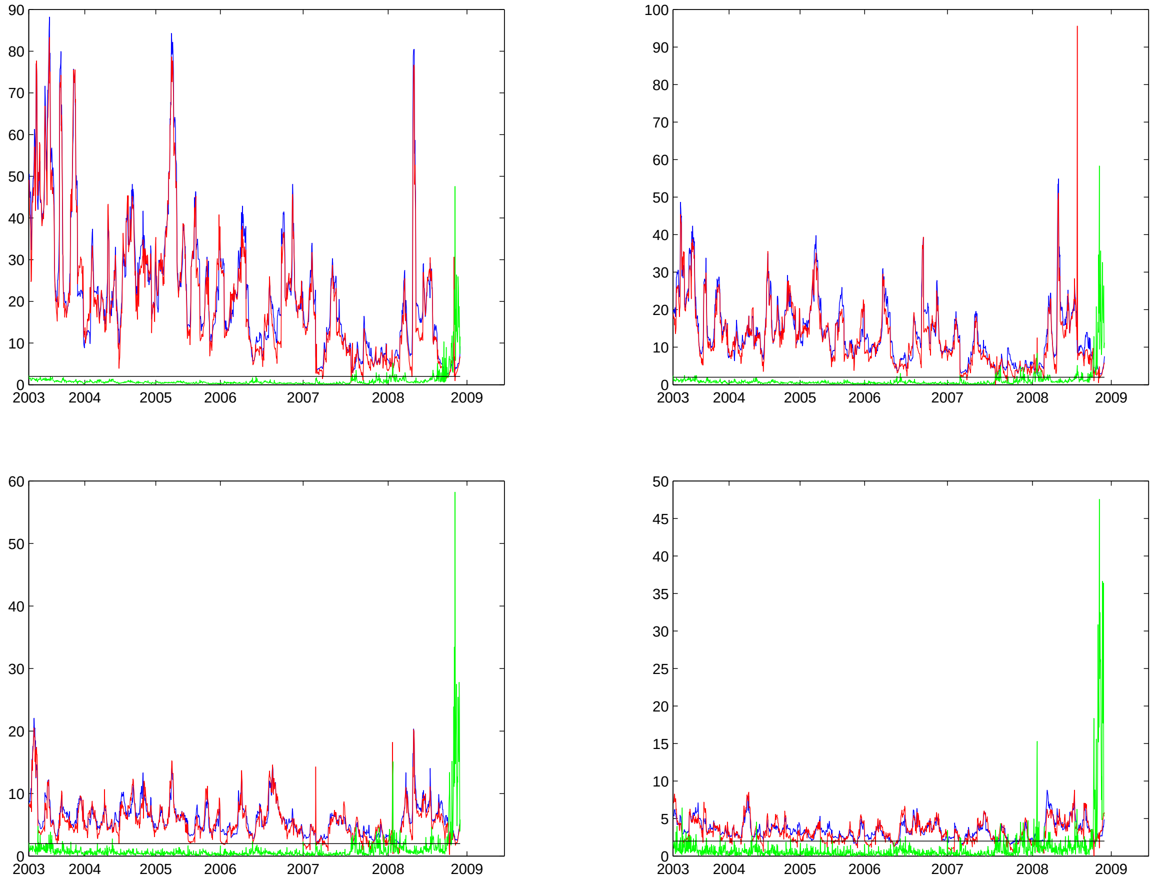


Figure 3.10: Evolution degrees of freedom estimated with K_B (blue line) and K_G (red line) when a rolling window of length 21 trading days is used. The sequence of realized covariances matrix was obtained using the Standard estimators at 15 seconds (top left), 5 minutes (top right), 30 minutes (bottom left) and 65 minutes (bottom right) sampling frequency. The green line at the bottom represents the realized volatility of the equally-weighted portfolio consisting of the S&P 500 and NASDAQ 100 futures. The black horizontal line coincides with the level $K = 2$. Values of the degrees of freedom below this line indicate the WAR model, in that particular window, is degenerate.

where C is a $n \times r$ matrix such that $C'B = \mathbf{0}_{(r,r)}$, i.e. C is orthogonal to B. Being B of dimension $n \times r$ and full rank, such a matrix C always exists. $z_{k,i,t}$ is the i-th element of the vector $z_{k,t} \equiv C'x_{k,t}$, of dimension $r \times 1$. $\hat{\Sigma}$ is an estimator of Σ. This estimator has an asymptotic distribution that is not normal and is given by

$$\hat{K} \xrightarrow{d} K \int_0^1 [W(s)]^2 ds, \tag{3.7.4}$$

where $W(\cdot)$ stands for the standard Brownian motion. Equation (3.7.4) implies that the estimator of the degrees of freedom derived under cointegration assumptions converges in distribution to a random variable with expectation strictly smaller than K. To conclude, a very low level of the estimated degrees of freedom for the WAR process might be due to the fact that the process is

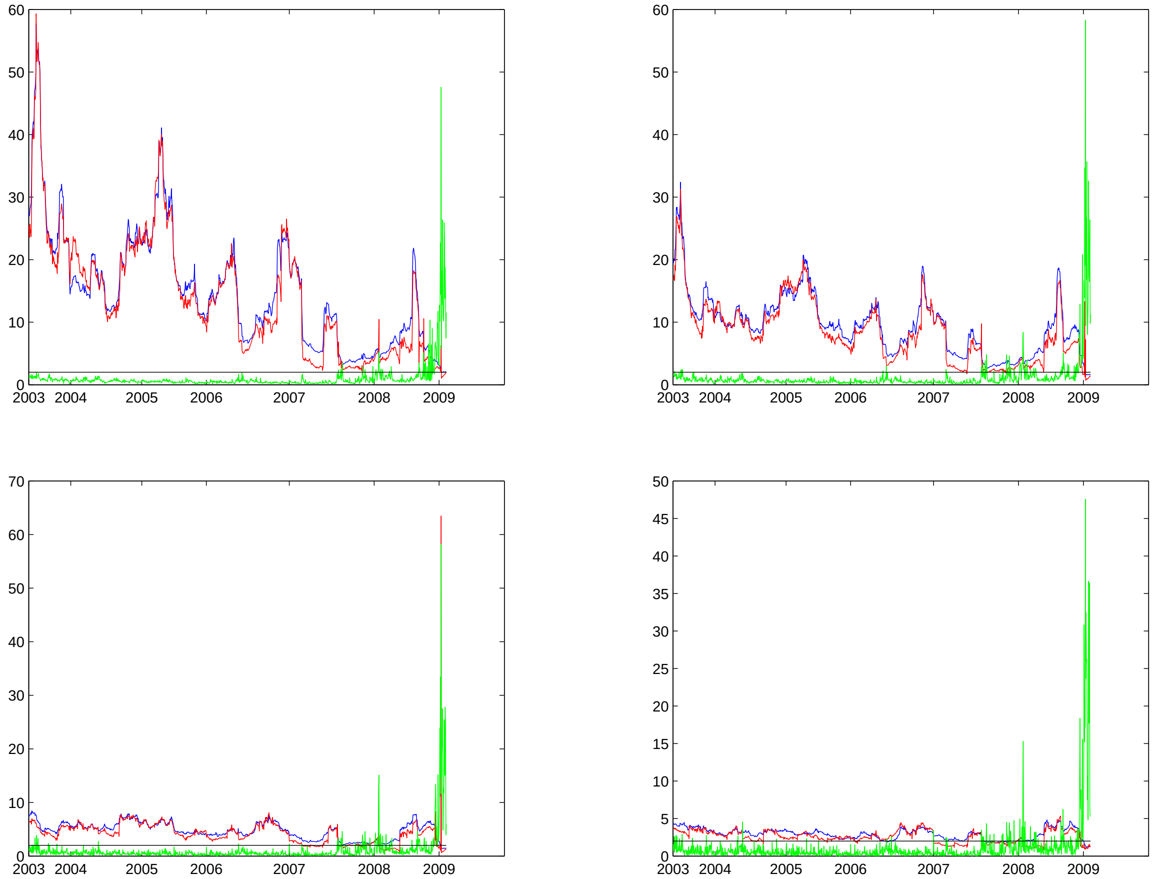


Figure 3.11: Evolution degrees of freedom estimated with K_B (blue line) and K_G (red line) when a rolling window of length 62 trading days is used. The sequence of realized covariances matrix was obtained using the Standard estimators at 15 seconds (top left), 5 minutes (top right), 30 minutes (bottom left) and 65 minutes (bottom right) sampling frequency. The green line at the bottom represents the realized volatility of the equally-weighted portfolio consisting of the S&P 500 and NASDAQ 100 futures. The black horizontal line coincides with the level $K = 2$. Values of the degrees of freedom below this line indicate the WAR model, in that particular window, is degenerate.

not stationary but is cointegrated.

In this section we investigate the plausibility of the assumption of cointegration for the WAR process. Following the lines of Chiriac (2007), we simulated a NoWAR(1) with autoregressive matrix and covariance matrix:

$$M = \begin{pmatrix} 0 & 0.8 \\ 0 & 1 \end{pmatrix} \ \Sigma = \begin{pmatrix} 0.08 & 0.09 \\ 0.09 & 0.20 \end{pmatrix}.$$

and $K = 3$. The rank of the matrix $H = \mathbf{2} - M$ is 1, thus each of the latent process are cointegrated with cointegration vector $\Gamma = (1, -0.8)$. The entries of Σ are the entries of the estimated matrix for the couple S&P 500 - NASDAQ 100 for the low volatility period. Figure 3.12 shows on the top the simulated realized volatility under the NoWAR assumption. Except for the different scale,

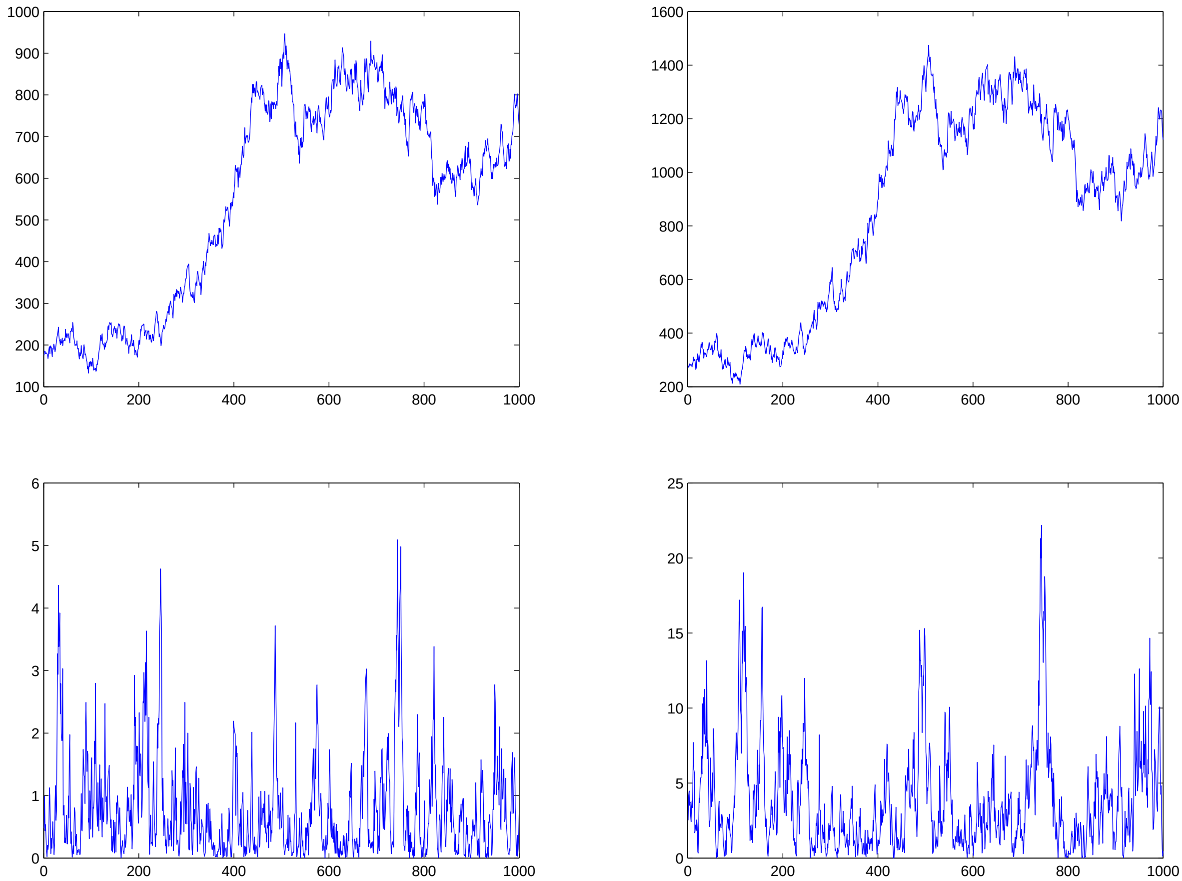


Figure 3.12: Simulated realized volatilities. Top panels: simulation under the cointegrated WAR process. Bottom panels: simulation under the stationary WAR model.

the two plots have a similar pattern that is more compatible with a price process rather than with the volatility of an asset. At the bottom, the plots represent the realized volatilities for simulated stationary WAR(1) process with the same covariance matrix Σ and autoregressive matrix M with roots close to 1,

$$M = \begin{pmatrix} 0.9 & 0 \\ 0 & 0.95 \end{pmatrix}$$

In this experiment we observe directly the lag of one time between the peaks and throughs as consequence of the recursive form of the matrix M. These two plots are more in line with the plots of realized volatility found in literature except for the higher peaks usually present (and missing here) that are caused by extreme market events.

3.8 Conclusions

In this paper we studied the estimation of the degrees of freedom K of a WAR process. The value of the degrees of freedom is important to determine whether the process possesses a density and if it is not degenerate. We introduced an alternative estimator of K, named K_B, and showed that, beside being more efficient, this novel estimator has lower MSE and MAD when compared to the standard estimator of Gourieroux et al. (2009), K_G. We also investigated a possible cause for very low values of the estimated degrees of freedom and we found that this might be explained by the presence of extreme events in the variance-covariance process. Finally, using high-frequency data, we estimated the degrees of freedom of the WAR model applied to the series of variance-covariance matrices of future prices of the pair S&P 500 - NASDAQ 100 indexes. We found that, no matter which estimator is used, the estimated degrees of freedom increases with the sampling frequency and do not depend on the variance of the realized covariance matrices estimators. We also suggested that the degrees of freedom are not likely to be constant over time and a rolling windows seems a more appropriate choice with respect to a single parameters estimation for the entire series of matrices. A more comprehensive empirical analysis on larger and different data set, along with a model that accounts for time-varying degrees of freedom, need to be investigated and are left for future research.

Appendix 3.A Estimation of K under stationarity assumption

This section reports part 4 of the Appendix in Chiriac (2007). Let $x_{k,t}$ with $k = 1, \ldots, K$ be independent VAR(1) processes of dimension $n \times 1$:

$$x_{k,t} = Mx_{k,t} + \epsilon_{k,t}, \qquad \epsilon k, t \overset{i.i.d.}{\sim} N(0, \Sigma). \tag{3.A.1}$$

Then the process $x_{k,t}$ has unconditional mean equal to 0 and unconditional variance given by:

$$V[x_{k,t+1}] = E[x_{k,t+1}x'_{k,t+1}] = MV[x_{k,t}]M' + V[\epsilon_{k,t+1}]. \tag{3.A.2}$$

Denote $V[x_{k,t}] \equiv \Sigma(\infty)$. From (3.A.2) it follows that:

$$\Sigma(\infty) = M\Sigma(\infty)M' + Sigma. \tag{3.A.3}$$

It results that $x_{k,t} \overset{i.i.d}{\sim} N(0, \Sigma(\infty))$ for $k = 1, \ldots, K$.

Let Y_t defined as in Equation (3.2.1):

$$Y_t = \sum_{k=1}^{K} x_{k,t}x'_{k,t}.$$

The unconditional mean of Y_t is equal to:

$$E[Y_t] = E\left[\sum_{k=1}^{K} x_{k,t}x'_{k,t}\right] = \sum_{k=1}^{K} E[x_{k,t}x'_{k,t}] = K\Sigma(\infty) \equiv \Sigma^*(\infty). \tag{3.A.4}$$

By multiplying Equation (3.A.3) by K, it follows that:

$$\Sigma^*(\infty) = M\Sigma^*(\infty)M' + \Sigma^*, \tag{3.A.5}$$

where $\Sigma^*(\infty) = K\Sigma(\infty)$.

Give a vector ω (portfolio allocation) of dimension $n \times 1$, the unconditional variance of $\omega Y_t \omega$

(portfolio volatility) is equal to:

$$\begin{aligned}
V[\omega' Y_t \omega] &= V[\omega' \sum_{i=1}^{K} x_{k,t} x'_{k,t} \omega] \\
&= \sum_{i=1}^{K} V[\omega' x_{k,t} x'_{k,t} \omega] \\
&= KV[\omega' x_{k,t} x'_{k,t} \omega] \\
&= KV[(\omega' x_{k,t})^2] \\
&= K\left(E[(\omega' x_{k,t})^4] - E[(\omega' x_{k,t})^2]^2\right) \\
&= 3K(\omega' V[x_{k,t}]\omega)^2 - K(\omega' V[x_{k,t}]\omega)^2 \quad \text{given that } \omega' x_{k,t} \overset{i.i.d.}{\sim} N(0, \omega'\Sigma(\infty)\omega) \\
&= 3K(\omega'\Sigma(\infty)\omega)^2 - K(\omega'\Sigma\omega)^2 \\
&= 2K(\omega'\Sigma(\infty)\omega)^2 \\
&= \frac{2}{K}(\omega'\Sigma(\infty)\omega)^2. \qquad (3.A.6)
\end{aligned}$$

therefore we conclude that

$$\hat{K} = \frac{2(\omega\Sigma^*(\infty))^2}{V[\omega' Y_t \omega]} \qquad (3.A.7)$$

3.A.1 Distribution K under stationarity assumption

This section report the paragraph 2.1 in Chiriac (2007).

The process defined in Equation (3.2.2) is strictly stationary, i.e. the WAR(1) process defined in Equation (3.2.1) is (strictly stationary), if and only if the matrix M has roots with modulus (strictly) less than 1. Under such conditions, we can write (3.2.2) as:

$$x_{k,t} = M^t x_{k,0} + \epsilon_{k,t} + M\epsilon_{k,t-1} + \cdots + M_{t-1}\epsilon_{k,0}. \qquad (3.A.8)$$

From the above representation, the distribution if $x_{k,t}$ conditional on $x_{k,0}$ for each $k = 1, \ldots, K$ is given by

$$x_{k,t}|x_{k,0} \overset{i.i.d.}{\sim} N(M^t x_{k,0}, \Sigma(*)), \qquad (3.A.9)$$

where

$$\Sigma(*) = \Sigma + M\Sigma M' + M^2\Sigma(M^2)^+ \cdots + M^{t-1}\Sigma(M^{t-1})'. \qquad (3.A.10)$$

The process Y_t is given by

$$\begin{aligned} Y_t &= \sum_{k=1}^{K} x_{k,t}x'_{k,t} = \sum_{k=1}^{K}(Mx_{k,t-1} + \epsilon_{k,t})(Mx_{k,t-1} + \epsilon_{k,t})' \\ &= \sum_{k=1}^{K}(Mx_{k,t-1}x'_{k,t-1}M' + Mx_{k,t-1}\epsilon'_{k,t} + \epsilon_{k,t}x'_{k,t-1}M' + \epsilon_{k,t}\epsilon'_{k,t}) \\ &= MY_{t-1}M' + \sum_{k=1}^{K} Mx_{k,t-1}\epsilon'_{k,t} + \sum_{k=1}^{K}\epsilon_{k,t}x'_{k,t-1}M' + \sum_{k=1}^{K}\epsilon_{k,t}\epsilon'_{k,t}, \end{aligned} \tag{3.A.11}$$

and its conditional mean with respect to $\mathcal{F}_{t-1}$, the information set at time $t-1$, is

$$E[Y_t|\mathcal{F}_{t-1}] = MY_{t-1}M' + K\Sigma. \tag{3.A.12}$$

The conditional expectation with respect to the information set at time $t = 0$ is given by:

$$\begin{aligned} E[Y_t|\mathcal{F}_0] &= MY_0M' + K\Sigma + KM\Sigma M' + KM^2\Sigma(M^2)' + \cdots + KM^{t-1}\Sigma(M^{t-1})' \\ &= MY_0M' + K\Sigma(*). \end{aligned} \tag{3.A.13}$$

Given the assumption that $x_{k,0} = 0$, it follows that $Y_0 = 0_{n,n}$. In order to solve for K in Equation (3.A.13), which is a scalar. and given that $E[Y_t|\mathcal{F}_0]$ and $\Sigma(*)$ are matrices of dimension $n \times n$, multiplying both sides of (3.A.13) by a vector α of dimension $n \times 1$, leads to

$$K = \frac{\alpha' E[Y_{t|\mathcal{F}_0}]\alpha}{\alpha'\Sigma(*)\alpha}. \tag{3.A.14}$$

A consistent estimator of K is given then by:

$$\hat{K} = \frac{\frac{\sum_{t=1}^{t}\alpha' Y_t\alpha}{T}}{\alpha'\Sigma(**)\alpha} = \frac{\sum_{t=1}^{t}\sum_{k=1}^{K}\alpha' x_{k,t}x'_{k,t}\alpha}{T\alpha\Sigma(**)\alpha}, \tag{3.A.15}$$

where

$$\Sigma(**) = \Sigma + M\Sigma M' + M^2\Sigma(M^2)' + \cdots + M^{T-1}\Sigma(M^{T-1})'. \tag{3.A.16}$$

The asymptotic distribution of the estimated Wishart degrees of freedom K, assuming stationarity, is a normal distribution with mean 0 and variance γ_j (see Appendix A.1 in Chiriac (2007) for the proof):

$$\sqrt{T}(\hat{K} - K) \overset{i.i.d.}{\sim} N(0, \sum_{j=-\infty}^{\infty}\gamma_j), \tag{3.A.17}$$

where $\gamma_j = E[(S_t - \mu)(S_{t-j} - \mu)]$ with $S_t = \sum_{k=1}^{K}\frac{(\alpha' x_{k,t})^2}{\alpha'\Sigma(**)\alpha}$ and $\mu \equiv ES_t = K$ for all t.

Appendix 3.B Estimation K for a general $WAR(p)$ model

Let $Y_t \in \mathbb{R}^n \times \mathbb{R}^n$ be a WAR(p) process:

$$E\left[Y_t | \mathbb{I}_{t-1}\right] = \sum_{j=1}^{p} M_j Y_{t-j} M_j' + K\Sigma. \tag{3.B.1}$$

where $\mathbb{I}_{t-1}$ is the information set available up to time $t-1$.

Under stationary conditions, the unconditional mean of the process, $E\left[Y_t\right]$ is obtained using the law of iterated expected values:

$$E\left[Y_t\right] = E\left[E\left[Y_t | \mathbb{I}_{t-1}\right]\right] = \sum_{j=1}^{p} M_j E\left[Y_{t-j}\right] M_j' + K\Sigma \tag{3.B.2}$$

As the unconditional distribution of any WAR(p) process is a centered Wishart distribution, applying the definition of centered Wishart distribution, we can write

$$Y_t = \sum_{k=1}^{K} z_{k,t} z_{k,t}', \tag{3.B.3}$$

where $z_{t,k} \overset{i.i.d.}{\sim} N(0, \Sigma(\infty))$.

From (3.B.3) we have that

$$\begin{aligned} E\left[Y_t\right] &= \sum_{k=1}^{K} E\left[z_{k,t} z_{k,t}'\right] \\ &= KV\left[z_{k,t}\right] \\ &= K\Sigma(\infty). \end{aligned} \tag{3.B.4}$$

Combining this result with (3.B.2) and defining $\Sigma^*(\infty) = K\Sigma(\infty)$ and $\Sigma^* = K\Sigma$ we get:

$$\begin{aligned} \Sigma^*(\infty) &= \sum_{j=1}^{p} M_j E\left[Y_{t-j}\right] M_j' + K\Sigma \\ &= \sum_{j=1}^{p} M_j K\Sigma(\infty) M_j' + K\Sigma \\ &= \sum_{j=1}^{p} M_j \Sigma^*(\infty) M_j' + \Sigma^* \end{aligned} \tag{3.B.5}$$

From (3.3.14) we know that, for any given vector $\omega \in \mathbb{R}^n$

$$\omega' Y_t \omega \sim \mathrm{Ga}(K/2, 2\omega'\Sigma(\infty)\omega). \tag{3.B.6}$$

Knowing the variance of a gamma distributed random variable we have

$$V\left[\omega' Y_t \omega\right] = \frac{K}{2}(2\omega'\Sigma(\infty)\omega)^2. \tag{3.B.7}$$

$\Sigma(\infty)$ is not observable, but given the estimated matrices $\hat{M}_j,\ j = 1, \ldots, p$ and $\hat{\Sigma}^*$ we can recover $\hat{\Sigma}^*(\infty)$ that satisfies (3.B.5). Thus:

$$\begin{aligned} V\left[\omega' Y_t \omega\right] &= \frac{K}{2}\left(2\omega'\frac{\hat{\Sigma}^*(\infty)}{K}\omega\right)^2 \\ &= \frac{2}{K}\left(\omega'\hat{\Sigma}^*(\infty)\omega\right)^2. \end{aligned} \tag{3.B.8}$$

Therefore the estimated degrees of freedom are

$$\hat{K} = \frac{2(\omega'\hat{\Sigma}^*(\infty)\omega)^2}{V\left[\omega' Y_t \omega\right]} \tag{3.B.9}$$

Appendix 3.C Modeling extreme events: proof of Proposition 2

Unconditional case:
$(i1) + (u2)$

As for any $Y \sim W(K, \Sigma)$, i.e. Y follows a centered Wishart distribution, we have that $\alpha' Y \alpha \sim Ga(K/2, 2\alpha'\Sigma\alpha)$. So that:

$$\begin{aligned} E[\alpha' Y \alpha] &= K\alpha'\Sigma\alpha \\ V[\alpha' Y \alpha] &= 2K(\alpha'\Sigma\alpha)^2 \\ E[(\alpha' Y \alpha)^2] &= K(K+2)(\alpha'\Sigma\alpha)^2 \end{aligned}$$

Then when the density of Y comes from a mixture of Wishart as in Equation 3.5.3 we have:

$$\begin{aligned} E[\alpha' Y \alpha] &= pK\alpha'\frac{\Sigma}{2p}\alpha + (1-p)K\alpha'\frac{\Sigma}{2)1-p)}\alpha \\ &= K\alpha'\Sigma\alpha \\ E[(\alpha' Y \alpha)^2] &= p\left(\alpha'\frac{\Sigma}{2p}\alpha\right)^2 K(K+2) + (1-p)\left(\alpha'\frac{\Sigma}{2(1-p)}\alpha\right)^2 K(K+2) \\ &= \frac{1}{4}K(K+2)(\alpha'\Sigma\alpha)^2\frac{1}{p(1-p)} \end{aligned}$$

this leads to

$$
\begin{aligned}
V[\alpha' Y \alpha] &= \frac{1}{4}K(K+2)(\alpha'\Sigma\alpha)^2 \frac{1}{p(1-p)} - K^2(\alpha'\Sigma\alpha)^2 \\
&= (\alpha'\Sigma\alpha)^2 K \left(\frac{K+2}{4p(1-p)} - K \right).
\end{aligned}
$$

$(u3)$

Trivially proven substituting $p = 1/2$.

$(u4)$

For a given $p \in]0,1[$, the Fourier transform of the p.d.f. (3.5.3), denoted $\phi^Y_{K,\Sigma}$ is given by:

$$
\begin{aligned}
\phi^Y_K(\Omega) &= E\left[\exp i \mathrm{Tr} Y \Omega\right] \\
&= p \left| \mathbf{I} - 2i\frac{\Sigma}{2p}\Omega \right|^{-K/2} + (1-p) \left| \mathbf{I} - 2i\frac{\Sigma}{2(1-p)}\Omega \right|^{-K/2}
\end{aligned}
$$

and in the extreme case event

$$
\lim_{p\to 0} \phi^Y_K(\Omega) \neq \left| \mathbf{I} - 2i\frac{\Sigma}{2}\Omega \right|^{-K/2}
$$

which is the Fourier transform of a Wishart distribution with degrees of freedom K and scale matrix $\Sigma/2$. Consequently, the sequence of matrices with p.d.f. 3.5.3 does not converges in distribution to a $W[K, \Sigma/2]$ when $p \to 0$. This result could also been see in (u2) from the face that when $p \to 0$, the variance of $\alpha' Y \alpha'$ explodes to infinite thus Y does not converges in distribution to a Wishart process.

Conditional case:

$(c1) + (c2)$

In the simple case where $Y_t \sim W[M, K, \Sigma]$, i.e. it follows a WAR(1) process we have that:

$$
\begin{aligned}
E[\alpha' Y_t \alpha | \mathbb{I}_{t-1}] &= \alpha' M Y_{t-1} M' \alpha + K \alpha' \Sigma \alpha \\
V[Y_t | \mathbb{I}_{t-1}] &= 4\alpha' M Y_{t-1} M' \alpha \alpha' \Sigma \alpha + 2K(\alpha'\Sigma\alpha)^2
\end{aligned}
$$

and so

$$
\begin{aligned}
E[(\alpha' Y_t \alpha)^2 | \mathbb{I}_{t-1}] &= V[Y_t | \mathbb{I}_{t-1}] + E[\alpha' Y_t \alpha | \mathbb{I}_{t-1}]^2 \\
&= 4\alpha' M Y_{t-1} M' \alpha \alpha' \Sigma \alpha + 2K(\alpha'\Sigma\alpha)^2 + (\alpha' M Y_{t-1} M' \alpha)^2 + \\
&\quad K^2(\alpha'\Sigma\alpha)^2 + 2K\alpha' M Y_{t-1} M' \alpha \alpha' \Sigma \alpha.
\end{aligned}
$$

When the p.d.f. of $Y_t | \mathbb{I}_{t-1}$ comes from a mixture of WAR(1) as in Equation 3.5.4 we have that:

$$\begin{aligned}
E[\alpha' Y_t \alpha | \mathbb{I}_{t-1}] &= \alpha' \left(p \left(MY_{t-1}M' + K\frac{\Sigma}{2p} \right) + (1-p) \left(MY_{t-1}M' + K\frac{\Sigma}{2(1-p)} \right) \right) \alpha \\
&= \alpha' MY_{t-1}M'\alpha + K\alpha'\Sigma\alpha.x \\
E[(\alpha' Y_t \alpha)^2 | \mathbb{I}_{t-1}] &= p \left(4\alpha' MY_{t-1}M'\alpha\alpha'\frac{\Sigma}{2p}\alpha + 2K(\alpha'\frac{\Sigma}{2p}\alpha)^2 + (\alpha' MY_{t-1}M'\alpha)^2 + K^2(\alpha'\frac{\Sigma}{2p}\alpha)^2 + \right. \\
&\quad \left. 2K\alpha' MY_{t-1}M'\alpha\alpha'\frac{\Sigma}{2p}\alpha \right) + (1-p) \left(4\alpha' MY_{t-1}M'\alpha\alpha'\frac{\Sigma}{2(1-p)}\alpha + \right. \\
&\quad 2K(\alpha'\frac{\Sigma}{2(1-p)}\alpha)^2 + (\alpha' MY_{t-1}M'\alpha)^2 + K^2(\alpha'\frac{\Sigma}{2(1-p)}\alpha)^2 + \\
&\quad \left. 2K\alpha' MY_{t-1}M'\alpha\alpha'\frac{\Sigma}{2(1-p)}\alpha \right) \\
&= 4\alpha' MY_{t-1}M'\alpha\alpha'\Sigma\alpha + (\alpha' MY_{t-1}M'\alpha^2 + \frac{K}{2}\frac{(\alpha'\Sigma\alpha)^2}{p} + \frac{K^2}{4}\frac{(\alpha'\Sigma\alpha)^2}{p} + \\
&\quad 2K\alpha' MY_{t-1}M'\alpha\alpha'\Sigma\alpha + \frac{K}{2}\frac{(\alpha'\Sigma\alpha)^2}{1-p} + \frac{K^2}{4}\frac{(\alpha'\Sigma\alpha)^2}{1-p} \\
&= 4\alpha' MY_{t-1}M'\alpha\alpha'\Sigma\alpha + (\alpha' MY_{t-1}M'\alpha^2 + 2K\alpha' MY_{t-1}M'\alpha\alpha'\Sigma\alpha + \\
&\quad \frac{K}{2}\frac{(\alpha'\Sigma\alpha)^2}{p(1-p)} + \frac{K^2}{4}\frac{(\alpha'\Sigma\alpha)^2}{p(1-p)}.
\end{aligned}$$

Now, given that

$$\begin{aligned}
E[\alpha' Y_t \alpha | \mathbb{I}_{t-1}]^2 &= (\alpha' MY_{t-1}M'\alpha + K\alpha'\Sigma\alpha)^2 \\
&= (\alpha' MY_{t-1}M'\alpha)^2 + K^2(\alpha'\Sigma\alpha)^2 + 2KMY_{t-1}M'\alpha\alpha'\Sigma\alpha,
\end{aligned}$$

we finally get

$$\begin{aligned}
V[\alpha' Y_t \alpha | \mathbb{I}_{t-1}] &= 4\alpha' MY_{t-1}M'\alpha\alpha'\Sigma\alpha + (\alpha' MY_{t-1}M'\alpha^2 + \\
&\quad 2K\alpha' MY_{t-1}M'\alpha\alpha'\Sigma\alpha + \frac{K}{2}\frac{(\alpha'\Sigma\alpha)^2}{p(1-p)} + \frac{K^2}{4}\frac{(\alpha'\Sigma\alpha)^2}{p(1-p)} \\
&\quad -(\alpha' MY_{t-1}M'\alpha)^2 - K^2(\alpha'\Sigma\alpha)^2 - 2KMY_{t-1}M'\alpha\alpha'\Sigma\alpha \\
&= 4\alpha' MY_{t-1}M'\alpha\alpha'\Sigma\alpha + \frac{K}{2}(\alpha'\Sigma\alpha)^2 \left[\frac{1}{p(p-1)}\frac{K}{2p(p-1)} - K \right].
\end{aligned}$$

$(c3)$
Same as $(u3)$.

$(c4)$
To prove this result we do not consider the Fourier transform (i.e. the characteristic function) but the Laplace transform, which is used to define a WAR process. The Laplace transform of the

p.d.f. 3.5.4 is (see Equation (3.2.3):

$$\begin{aligned}\mathbf{\Psi}_t(\Gamma) &= E[\exp Tr(\Gamma Y_t)] \\ &= p\frac{\exp\left(Tr[M'\Gamma(\mathbf{I}_d - 2\frac{\Sigma}{2p}\Gamma)^{-1}MY_{t-1}]\right)}{[\det(\mathbf{I}_d - 2\frac{\Sigma}{2p}\Gamma)]^{K/2}} + (1-p)\frac{\exp\left[Tr(M'\Gamma(\mathbf{I}_d - 2\frac{\Sigma}{2(1-p)}\Gamma)^{-1}MY_{t-1})\right]}{[\det(\mathbf{I}_d - 2\frac{\Sigma}{2(1-p)}\Gamma)]^{K/2}}.\end{aligned}$$

The Laplace transform is defined for a matrix Γ such that $\|2\frac{\Sigma}{2p}\Gamma\| < 1$ and $\|2\frac{\Sigma}{2(1-p)}\Gamma\| < 1$.

In the extreme risk case we have that

$$\lim_{p\to 0}\mathbf{\Psi}_t(\Gamma) \neq \frac{\exp\left(Tr[M'\Gamma(\mathbf{I}_d - 2\frac{\Sigma}{2}\Gamma)^{-1}MY_{t-1}]\right)}{[\det(\mathbf{I}_d - 2\frac{\Sigma}{2}\Gamma)]^{K/2}}$$

and thus $Y_t|\mathbb{I}_{t-1}$ does not converge in distribution to a $W[M, K, \Sigma/2]$ process. Again, as for (u4), one can see from (c2) that when $p \to 0$ the conditional variance of $\alpha' Y_t \alpha$ explodes and thus Y_t does not converge in distribution to a WAR process.

Appendix 3.D Simulation of from a stable Paretian distribution

Sub-Gaussian random vectors represent a special case of symmetric table random. Unlike the general case, this class of vectors possesses a tractable expression for the characteristic function and thus the estimation of the multivariate density is relatively easy. See Samorodnitsky and Taqqu (1994) for a complete exposition on the stable Paretian distribution.

The procedure to simulate sub-Gaussian random vectors with stability index α, denoted $S\alpha S$, is fairly easy for the fact that every $S\alpha S$ random variable has is conditionally Gaussian distribution. In fact, it can be shown that, taking any random variable A so that

$$A \sim S_{\alpha/2}\left(2(\cos\frac{\pi\alpha}{4})^{2/\alpha}, 1, 0\right) \tag{3.D.1}$$

with $\alpha < 2$ and a zero mean Gaussian vector in $\mathbb{R}^d$ $\mathbf{G} = (G_1, \ldots, G_d)$ independent of A we have that the random vector

$$\mathbf{X} = (A^{1/2}G_1, \ldots, A^{1/2}G_d) \tag{3.D.2}$$

has a $S\alpha S$ distribution in $\mathbb{R}^d$ because, for any real numbers $b_1, \ldots, b_d$ the linear combination $\sum_{k=1}^d A^{1/2}G_k = A^{1/2}\sum_{k=1}^d G_k$ is a $S\alpha S$ random variable and hence $\mathbf{X}$ is $S\alpha S$ and we write $\mathbf{X} \sim S_\alpha(\Sigma, 0, 0)$. (see Theorem 2.1.5 in for Samorodnitsky and Taqqu, 1994).

To simulate $\mathbf{X}$ we only need to simulate from a multivariate Gaussian distribution density and from an univariate stable density. To simulate from an univariate stable density one can use, for instance, the Chambers et al. (1976) method.

3.D.1 Figures

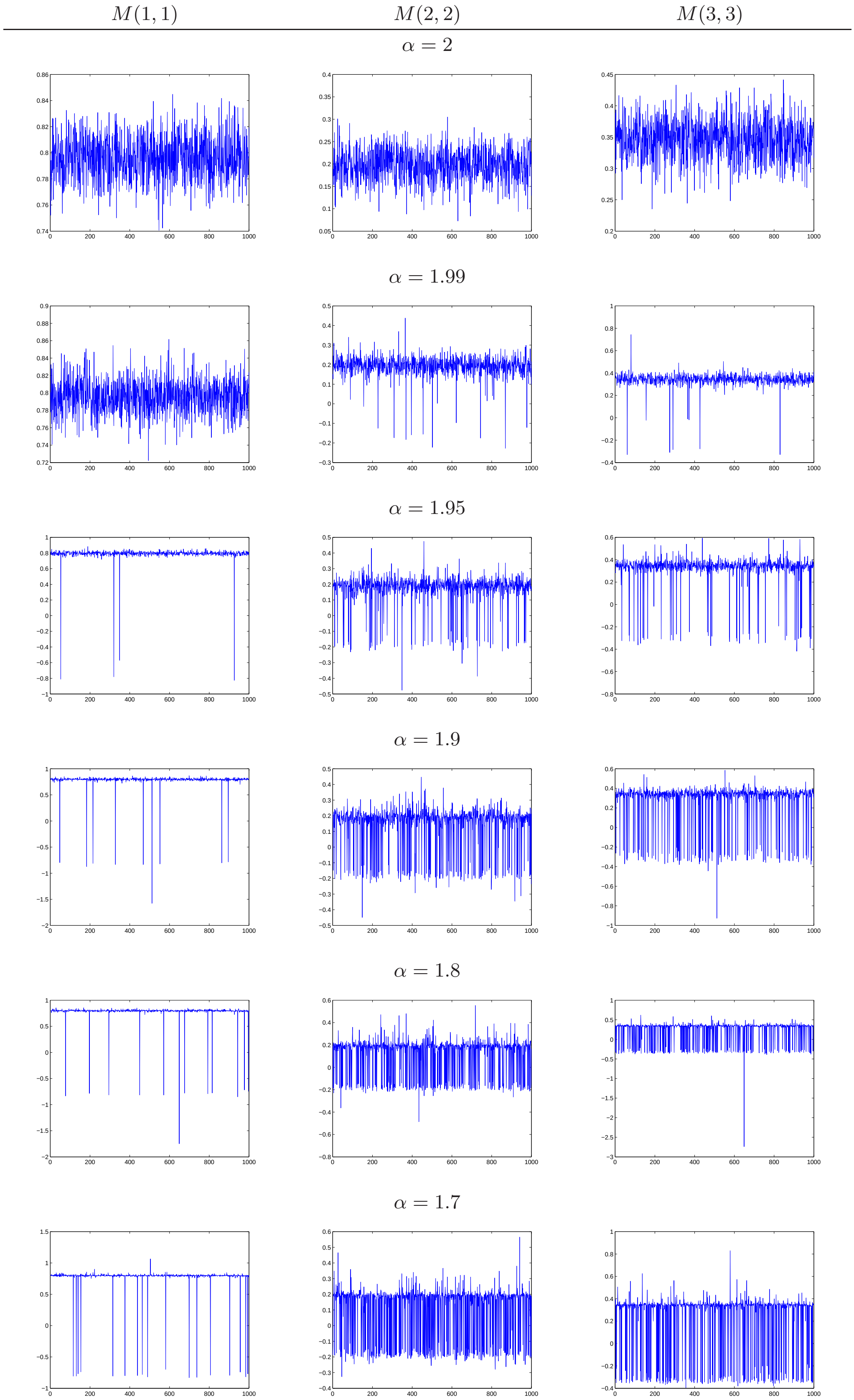


Figure 3.13: Estimated diagonal entries of M for the simulated misspecified WAR using stable Paretian random vectors.

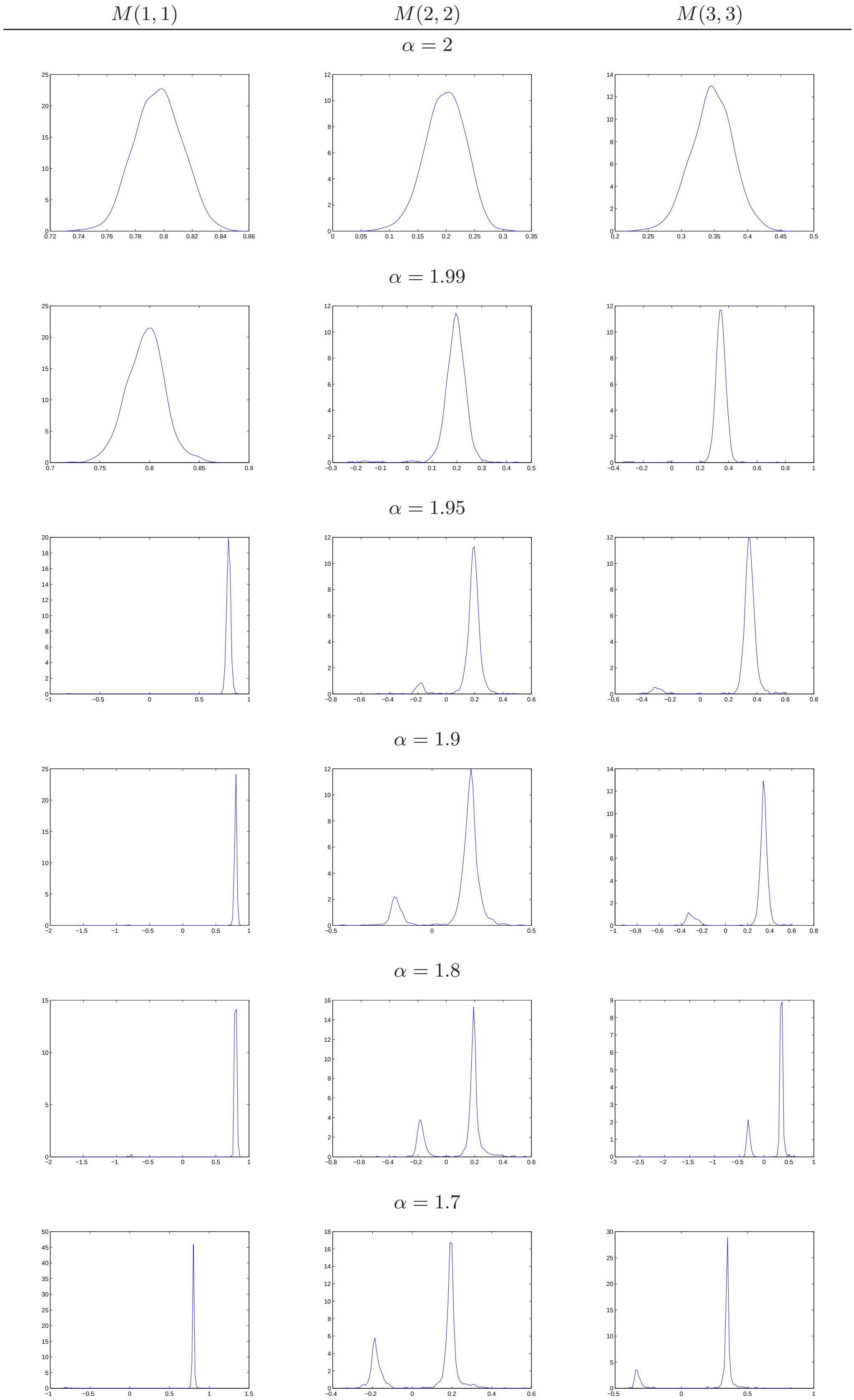


Figure 3.14: Kernel density of the estimated diagonal entries of M for the simulated misspecified WAR using stable Paretian random vectors.

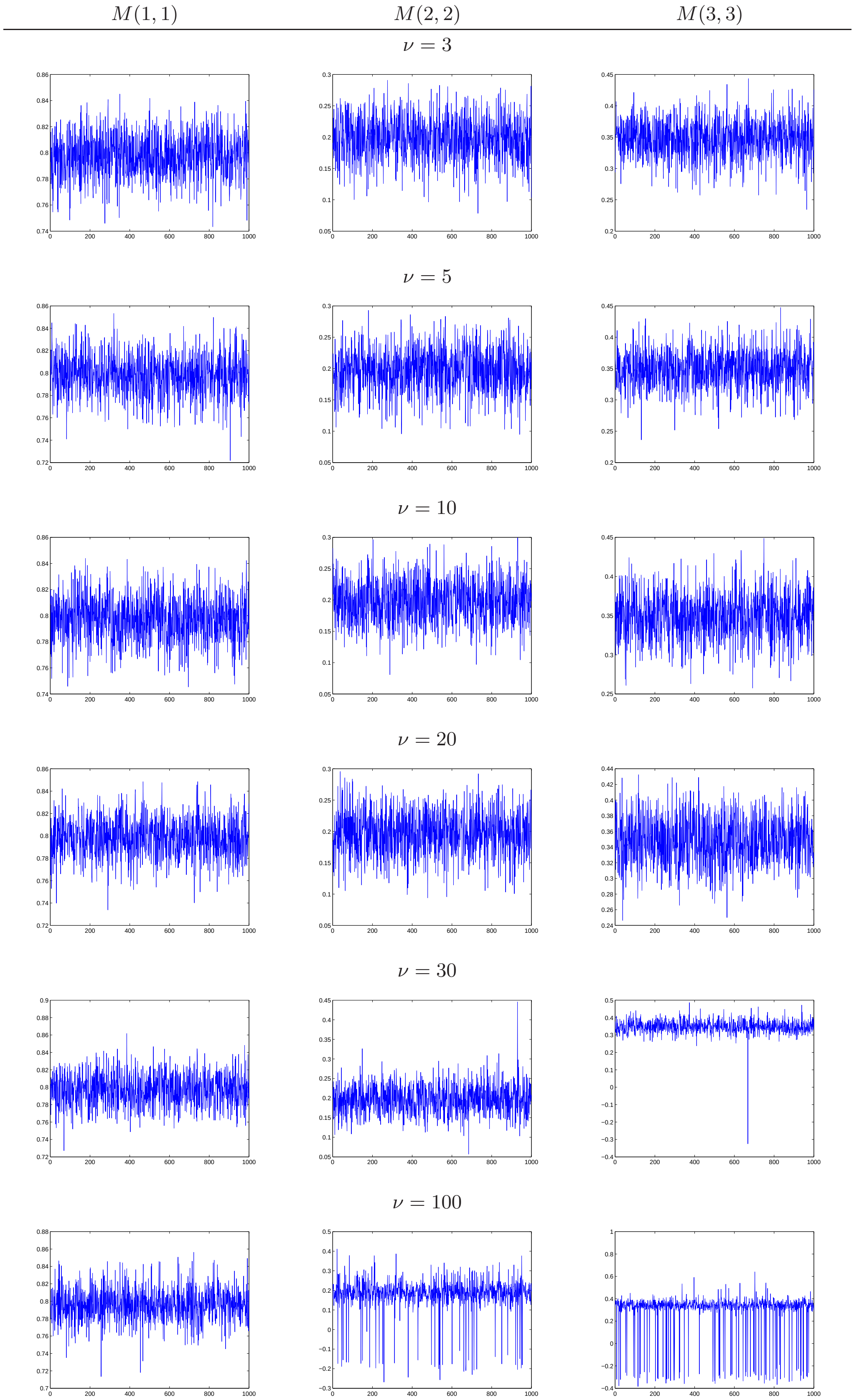


Figure 3.15: Estimated diagonal entries of M for the simulated misspecified WAR using Student's t random vectors.

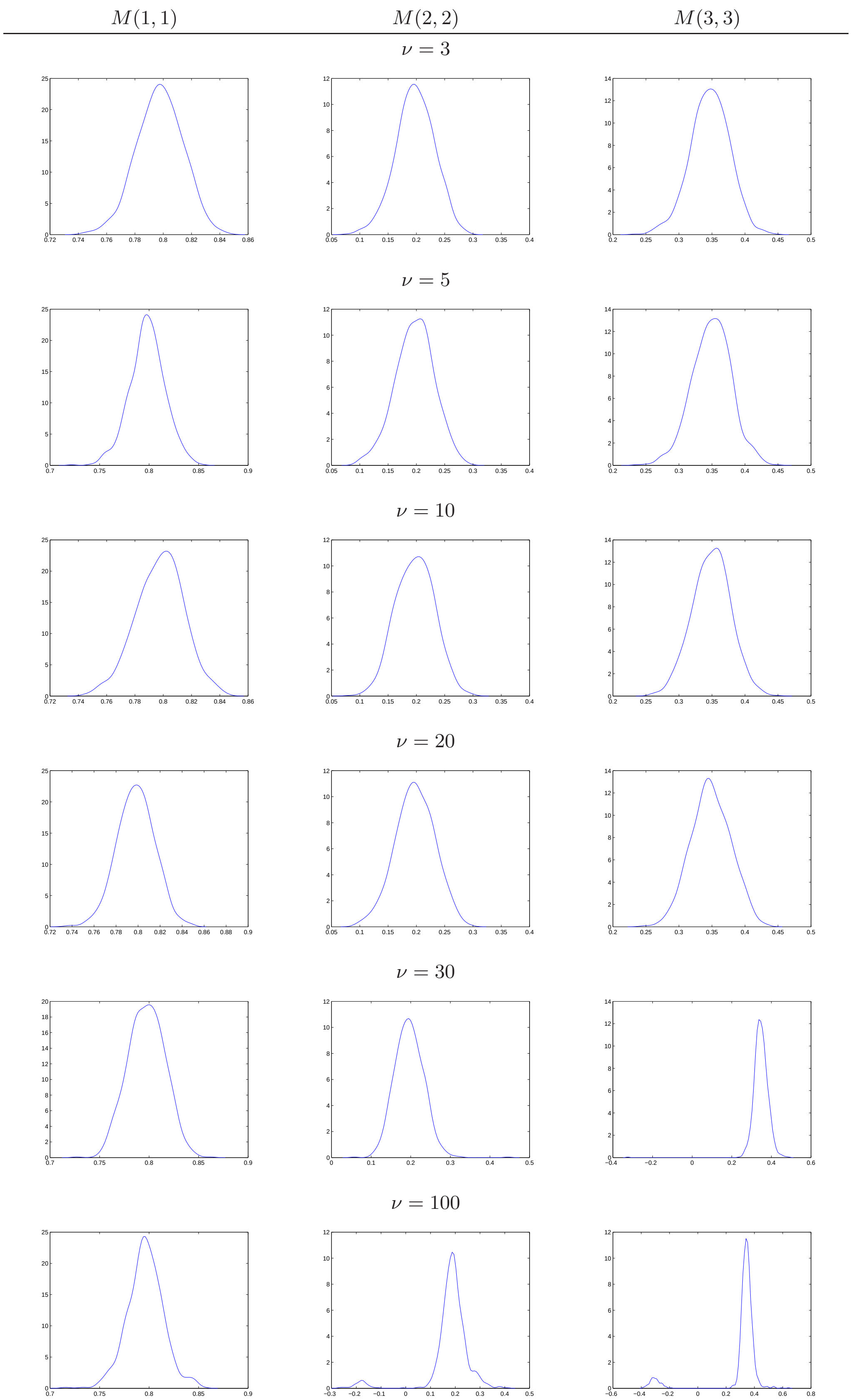


Figure 3.16: Kernel density of the estimated diagonal entries of M for the simulated misspecified WAR using Student's t random vectors.

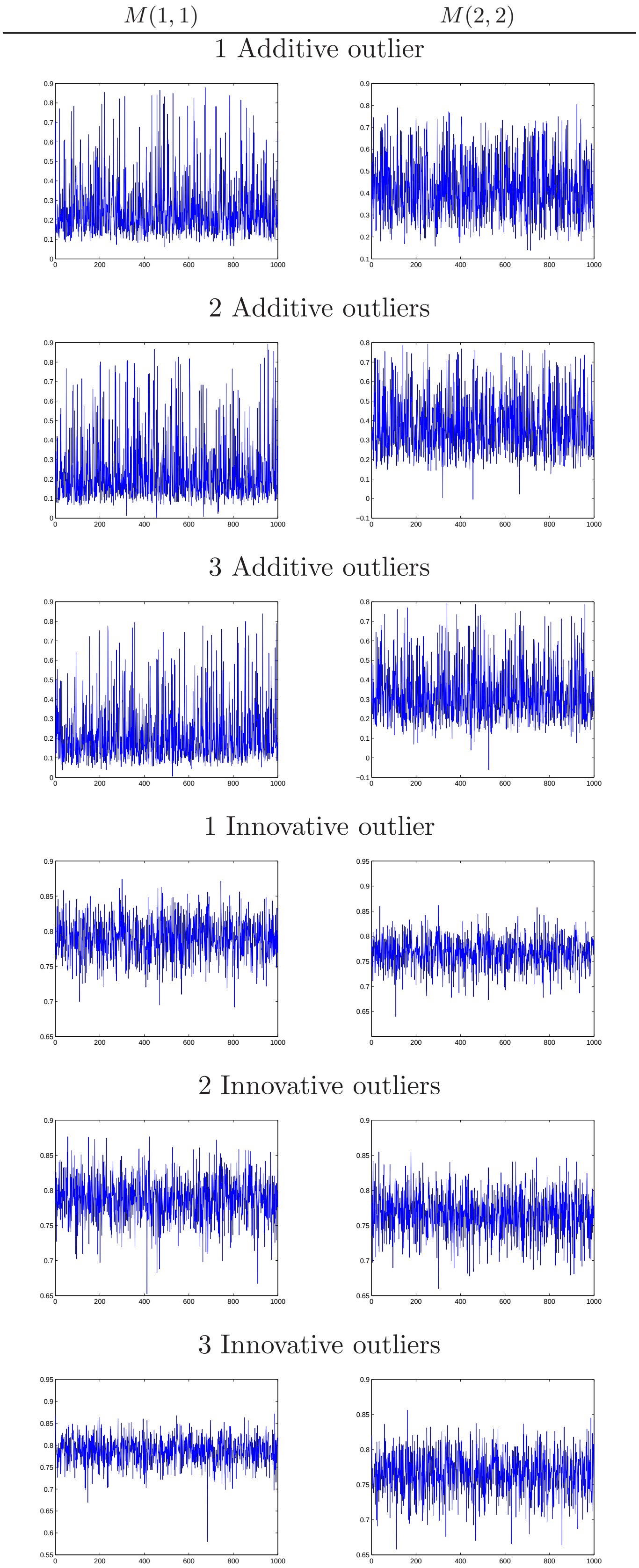


Figure 3.17: Estimated diagonal entries of of M when additive or innovative outliers are introduced.

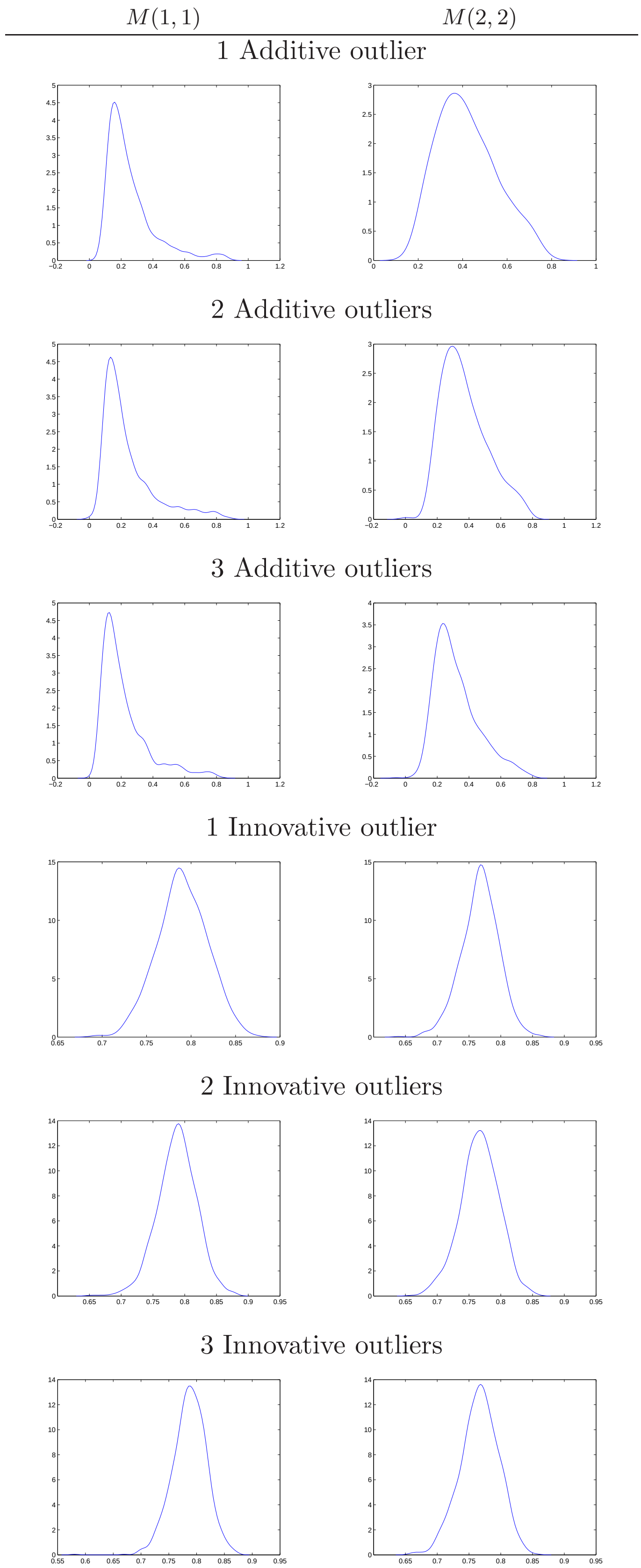


Figure 3.18: Kernel density of the estimated diagonal entries M when additive or innovative outliers are introduced.

Bibliography

Aas, K. and I. Haff (2006): "The generalized hyperbolic skew Student's t distribution," *Journal of Financial Econometrics*, 4, 275–309. 95

Adesi, G. B., R. Engle, and L. Mancini (2008): "A GARCH option pricing model with filtered historical simulation," *Review of Economic Studies*, 21, 1223–1258. vii

Adler, R. (1997): "Discussion: Heavy tail modeling and teletraffic data," *Annals of Statistics*, 25, 1849–1852. 5

Aït-Sahalia, Y. and L. Mancini (2008): "Out of sample forecast of quadratic variation," *Journal of Econometrics, forthcoming*. 51, 66, 89, 92

Akgiray, V. and G. Booth (1988): "The stable-law model of stock returns," *Journal of Business and Economic Statistics*, 6, 51–57. 16

Akgiray, V., G. Booth, and O. Loistl (1989): "Stable laws are inappropriate to for describing German stock returns," *Allgemeines Statistiches Archiv*, 73, 115–121. 16

Andersen, T. and T. Bollerslev (1997): "Heterogeneous information arrivals and return volatiliy dynamics: Uncovering the long run in high frequency data," *The Journal of Finance*, 52, 975–1005. 51

——— (1998): "Answering the skeptics: Yes, standard volatility models do provide accurate forecasts," *International Economic Review*, 39, 885–905. vii, 42, 78

Andersen, T., T. Bollerslev, F. Diebold, and H. Ebens (2001a): "The distribution of realized stock return volatility," *Journal of Financial Economics*, 43–76. vii, 51, 92

Andersen, T., T. Bollerslev, F. Diebold, and P. Labys (2001b): "The distribution of realized exchange rate volatility," *Journal of the American Statistical Association*, 96, 42–55. 51

——— (2003): "Modeling and forecasting realized volatility," *Econometrica*, 71, 579–625. 58, 66, 68, 69, 94, 101, 108

Andreou, E. and E. Ghysels (2002): "Rolling sample volatility estimators: some new theoretical, simulation and empirical results," *Journal of Business and Economic Statistics*, 20, 363–376. 42

Asai, M., M. Caporin, and M. McAleer (2008): "Block structure multivariate stochastic volatility," *Manuscript*. 44

Bandi, F. and J. Russel (2005): "Realized covariation, realized beta and microstructure noise," *unpublished paper.* 42, 78

Bandi, F., J. Russel, and C. Yang (2007): "Realized volatility forecasting and option pricing," *Journal of Econometrics*, 147, 34–46. viii

Bandi, F., J. Russel, and Y. Zhu (2006): "Using high-frequency data in dynamic portfolio choice," *Econometric Reviews, forthcoming.* viii, 43

Barndorff-Nielsen, O., P. Hansen, A. Lunde, and N. Shephard (2008a): "Designing Realized Kernels in to Measure the Ex-Post Variation of Equity Prices in the Presence of Noise," *Econometrica*, 76, 14811536. viii

——— (2008b): "Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading," *manuscript.* 59

Barndorff-Nielsen, O. and N. Shephard (2004): "Econometric analysis of realized covariation: high-frequency based covariance, regressions, and correlation in financial economics," *Econometrica*, 72, 885–925. 42, 58, 78

Battaglia, F. (2005): "Outliers is funcional autoregressive time series," *Statistics and Probability Letters*, 72, 323–332. 98, 102

Bauer, G. and K. Vorkink (2007): "Multivariate realized stock market volatility," *Bank of Canada Working Papers 2007-20.* 43

Bauwens, L. and S. Laurent (2005): "A new class of multivariate skew densities, with application to GARCH models," *Journal of Business and Economic Statistics*, 23, 346–354. 4, 22, 25, 26, 95

Bauwens, L., S. Laurent, and J. Rombouts (2006): "Multivariate GARCH models: a survey," *Journal of Applied Econometrics*, 21, 79–109. 22

Bera, A. and C. Jarque (1980): "Efficient tests for normality, homoscedasticity and serial independence of regression residuals," *Economic Letter*, 6, 255–259. 86

Berkowitz, J. (2001): "Testing densities forecasts, with applications to risk management," *Journal of Business and Economic Statistics*, 19, 465–474. 70, 73

Bertholon, H., A. Monfort, and F. Pegoraro (2009): "Pricing and Inference with Mixtures of Conditionally Normal Processes," *manuscript.* 104, 106, 107

Billio, M. and M. Caporin (2008): "A generalised dynamic conditional correlation model for portfolio risk evaluation," *Mathematics and Computers in Simulations, forthcoming.* 44

Billio, M., M. Caporin, and M. Gobbo (2006): "Flexible dynamic conditional correlation multivariate GARCH for asset allocation," *Applied Financial Economics Letters*, 2, 123–130. 44

Black, F. (1976): "Studies of stock market volatility changes," *Proceedings of the American Statistical Assiciation, Business and Economic Statistics Section*, 177–181. 16

Bollerslev, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327. vii

——— (1987): "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return," *The Review of Economics and Statistics and Probability Letters*, 69, 542–574. 95

Bonato, M. (2009a): "Estimating the degrees of freedom of the realized volatility Wishart autoregressive model," *Manuscript.* 56, 65

——— (2009b): "Modeling fat tails in stock returns: a multivariate Stable-GARCH approach," *manuscript.* 95

Bonato, M., M. Caporin, and A. Ranaldo (2008): "Forecasting realized (co)variances with a block structure Wishart autoregressive model," *manuscript.* 78, 82, 87, 108

Bradley, B. and M. Taqqu (2001): "Financial risk and heavy tails," in *Heavy tailed distributions in finance. Rachev S.T. editor*, North Holland. 4, 95

Brownlees, C. and G. Gallo (2008): "Comparison of Volatility Measures: A Risk Management Perspective," . viii, 69, 92

Cai, Y. and N. Davies (2003): "A simple diagnostic method of outlier detection for stationary Gaussian time series," *Journal of Applied Statistics*, 30, 205–223. 98

Caporin, M. and P. Paruolo (2008): "Spatial dependence in multivariate volatility models," *Working Paper.* 44

Chaboud, A., B. Chiquoine, E. Hjalmarsson, and C. Vega (2009): "Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market," *Board of Governors of the Federal Reserve System, mimeo.* 44

Chambers, J., C. Mallows, and B. Stuck (1976): "A method for simulating stable random variables," *Journal of the American Statistical Association*, 71, 340–344. 25, 124

Chen, C. and L. Liu (1993): "Forecasting time series with outliers," *Journal of Forecasting*, 12, 13–35. 98

Chernick, M., D. Downing, and D. Pike (1982): "Detecting outliers in time series data," *Journal of the American Statistical Association*, 77, 743–747. 98

Chiriac, R. (2006): "Estimating realized volatility Wishart autoregressive model," *Working Paper.* 79, 86, 97

——— (2007): "Nonstationary Wishart autoregressive model," *Working Paper.* 75, 79, 80, 85, 97, 112, 114, 117, 118, 119

Chiriac, R. and V. Voev (2008): "Modelling and forecasting multivariate realized volatility," *Working Paper.* 43

Christensen, K. and M. Podolskij (2007): "Realized range-based estimation of integrated variance," *Journal of Econometrics*, 323–349. viii

Clements, M. P., A. B. Galvao, and J. H. Kim (2008): "Quantile forecasts of daily exchange rates returns from forecasts of realized volatility," *Journal of Empirical Finance*, 15, 729–750. viii, 69

Corsi, F. (2009): "A Simple Approximate Long-Memory Model of Realized Volatility," *Journal of Financial Econometrics, forthcoming*, 7, 174–196. ix, 44, 51, 52

Corsi, F., U. Kretschmer, S. Mittnik, and C. Pigorsch (2007): "The volatility of realized volatility," *Journal of Financial Econometrics*, 27, 46–78. 51, 52

de Pooter, M., M. Martens, and D. van Dijk (2006): "Predicting the daily covariance matrix for S&P 100 stocks using intraday data - But which frequency to use?" *Econometric Reviews, forthcoming.* viii, 42, 43, 57, 58, 59, 89, 101, 108, 109

Doganoglu, T., C. Hartz, and S. Mittnik (2007): "Portfolio optimization when risk factors are conditionally varying and heavy tailed," *Computational Economics*, 29. 6, 7

Doganoglu, T. and S. Mittnik (2006): "Portfolio selection, risk assessment and infinite variance," *Working Paper.* 6, 7, 95

Engle, R. (1982): "Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kindom inflation," *Econometrica*, 11, 122–150. vii, 11

——— (2002): "Dynamic conditional correlation - a simple class of multivariate GARCH models," *Journal of Business and Economic Statistics*, 20, 339–350. 7, 22, 23, 24, 25, 27, 54

——— (2007): "High dimension dynamic correlation," *manuscript.* 7, 26, 27, 28

Engle, R. and G. Gallo (2006): "A Multiple Indicators Model for Volatility Using Intra-Daily Data," *Journal of Econometrics*, 131, 3–27. 68

Engle, R., D. Hendry, and J. Richard (1983): "Exogeneity," *Econometrica*, 51, 277–304. 30

Engle, R. and B. Kelly (2008): "Dynamic Equicorrelations," *Working Paper.* 44

Engle, R., N. Shephard, and K. Sheppard (2008): "Fitting vast dimensional time-varying covariance models," *manuscript.* 7, 27, 28, 29, 30

Engle, R. and K. Sheppard (2005): "Evaluating the specification of covariance models for large portfolios," *manuscript.* 27

Engle, R. F. and F. K. Kroner (1995): "Multivarariate simultaneous generalized ARCH," *Econometric Theory*, 11, 122–150. 64

Fama, E. (1965a): "The behavior of stock market prices," *Journal of Business*, 38, 34–105. 4, 95

——— (1965b): "Portfolio analysis in a stable Paretian market," *Management Science*, 11, 404–419. 4, 95

Feller, W. (1971): *An introducion to probabiliy theory and its applications*, John Wiley and Sons. 4

Fleming, J., C. Kirby, and B. Ostdiek (2001): "The economic value of volatility," *Journal of Finance*, 56, 329–352. 42

——— (2003): "The economic value of volatility timing using 'realized' volatiltiy," *Journal of Financial Economics*, 67, 473–509. 42

Foster, D. and D. Nelson (1996): "Continuous record asymptotics for rolling sample variance estimators," *Econometrica*, 64, 139–174. 42

Fox, A. (1972): "Outliers in time series," *Journal of the Royal Statistical Society - Series B*, 43, 350–363. 98

Gallant, R., C. Hsu, and G. Tauchen (1999): "Using daily range data to calibrate volatitliy diffusions and extract the forward integrated variance," *The Review of Economics and Statistics*, 81, 617–631. 51

Garcia, R., E. Renault, and D. Veredas (2006): "Estimation of stable distribution by indirect inference," *mimeo.* 6

Giot, P. and S. Laurent (2003): "Value-at-Risk for long and short trading positions," *Journal of Applied Econometrics*, 18, 641–664. 26, 95

——— (2004): "Modeling daily Value-at-Risk using realized volatiltiy and ARCH type models," *Journal of Empirical Finance*, 11, 379–398. viii, 69, 95

Gourieroux, C. (2007): "Positivity conditions for a bivariate autoregressive volatility specifications," *Journal of Financial Econometrics*, 5, 624–636. 46, 64, 65

Gourieroux, C., J. Jasiak, and R. Sufana (2009): "The Wishart autoregressive process of multivariate stochastic volatility," *Journal of Econometrics*, 150, 167–181, forthcoming. viii, 44, 45, 46, 53, 54, 55, 66, 69, 75, 78, 79, 80, 81, 82, 83, 85, 108, 116

Gourieroux, C., A. Monfort, and R. Sufana (2004): "International money and stock market contingent claims," *DP University of Toronto.* 81

Gourieroux, C. and J. Sufana (2007): "Nonlinear causality, with applications to liquidity and stochastic volatility," *manuscript.* 47, 78

Gourieroux, C. and R. Sufana (2003): "Wishart quadratic term structure models," *CREF 3-10, HEC Montreal.* 81

——— (2004): "Derivative pricing with multivariate stochastic volatility," *CREST DP.* 81

Granger, C. (1980): "Long memory relationships and the aggregation of dynamic models," *Journal of Econometrics*, 14, 227–238. 51

Haas, M., S. Mittnik, M. Paolella, and S. Steude (2005): "Stable mixture GARCH models," *MIMEO.* 17, 95

Hansen, P. and A. Lunde (2005): "A realized variance for the whole day based on intermittent high-frequency data," *Journal of Financial Econometrics*, 3, 525–554. 109

——— (2006): "Realized variance and market microstructure noise," *Journal of Business and Economic Statistics*, 24, 127–218. 109

Hayashi, T. and N. Yoshida (2005): "On covariance estimation of non-synchronously observed diffusion processes," *Bernoulli*, 11, 359–379. 42, 78

——— (2006): "Estimating correlations with nonsynchronous observations in continuous diffusion models," *Working Paper*. 42, 78

Hill, B. (1975): "A simple general approach to inference about the tail of a distribution," *Annals of Statistics*, 3, 1163–1174. 4, 34

Jasiak, J. and L. Lu (2007): "Causality and volatility transmission," *manuscript*. 47, 78

Kearns, P. and A. Pagan (1997): "Estimating the density tail index for financial time series," *The Review of Economics and Statistics*, 97, 171–175. 5

Keuster, K., S. Mittnik, and M. Paolella (2005): "Value-at-Risk Prediction: a Comparison of Alternative Strategies," *Journal of Financial Econometrics*, 4, 53–89. vii

Kim, T. and H. White (2004): "On more robust estimation of skewness and kurtosis," *Finance Research Letters*, 1, 65–70. 14, 94, 101

Kratz, M. and S. Resnick (1996): "The QQ-estimator and heavy tails," *Communication in Statistics - Stochastic Models*, 12, 699–724. 5

Kupiec, P. (1995): "Tecnique to verify the accuracy of risk measurament models," *Journal of Derivatives*, 2, 174–184. 26

Lamantia, F., S. Ortobelli, and S. Rachev (2005): "VaR, CVaR and time rules with elliptical and asymmetric stable distributed returns," *Working Paper*. 6

LeBaron, B. (2001): "Stochastic volatility as a simple generator of financial power-laws and long memory," *Quantitative Finance*, 1, 621 – 631. 51

Lombardi, M. and D. Veredas (2009): "Indirect estimation of elliptical stable distributions," *CORE discussion paper*, 53, 2309–2324. 6, 7, 32

Loretan, M. and P. Phillips (1994): "Testing the covariance stationarity of heavy tailed time series," *Journal of Empirical Finance*, 1, 211–285. 4

Lunde, A. and V. Voev (2007): "Integrated Covariance Estimation Using High-Frequency Data in the Presence of Noise," *Journal of Finacial Econometrics*, 5, 68–104. 59

Mancino, M. and S. Sanfelici (2008): "Estimating Covariance Via Fourier Method in the Presence of Asynchronous Trading and Microstructure Noise," *manuscript*. 59

Mandelbrot, B. (1963): "The variation of certain speculative prices," *Journal of Business*, 36, 391–419. 4, 95

Martens, M. and D. van Dijk (2007): "Measuring volatility with the realized range," *Journal of Econometrics*, 181–207. viii, 57, 89

McAleer, M. and M. Medeiros (2006): "Realized volatility: a review," *Working Paper*. 42, 78

McCulloch, J. (1986): "Simple consistent estimator of stable distribuition parameters," *Communication in Statistics - Simulation and Computation*, 15, 1109–1136. 35

——— (1997): "Measuring tail thckness in order to estomate the stable index α: a critique," *Journal of Business and Economic Statistics*, 15, 74–81. 5

——— (2003): "The risk neutral measure and option pricing under log-Stable uncertainty," *mimeo*. 4, 95

Meucci, A. (2005): *Risk and asset allocation*, Springer. 56, 74, 84, 85

Mittnik, S., T. Doganoglu, and D. Chenyao (1999a): "Computing the probability density funcion of the stable Paretian distributions," *Mathematical and Computing Modelling*, 29, 235–240. 5, 7, 18, 19, 32

Mittnik, S. and M. Paolella (1999): "A simple estimator for the characteristic exponent of the stable Paretian distribution," *Mathematical and Computer Modelling*, 29, 161–176. 34

Mittnik, S., M. Paolella, and S. Rachev (1998): "A tail estimator for the index of the stable Paretian distribution," *Communication in Statistics - Theory and Method*, 27. 5

——— (2000): "Diagnosing and treating the fat tails in financial returns data," *Journal of Empirical Finance*, 7, 389–416. 16, 95

——— (2002): "Stationarity of the stable power GARCH process," *Journal of Econometrics*, 106, 97–107. 17, 95

Mittnik, S. and S. Rachev (1993): "Reply to comments in 'Modeling asset returns with alternative stable models," *Econometric Reviews*, 12, 347–389. 5

Mittnik, S., S. Rachev, T. Doganoglu, and D. Chenyao (1999b): "Maximum likelihood estimation of stable Paretian models," *Mathematical and Computing Modelling*, 29, 275–293. 13, 20

Mouchel, W. D. (1973): "On the asymptotic normality of the maximu likelihood estimate when sampling from a stable distribution," *The Annals of Statistics*, 1, 948–957. 20

Muirhead, C. (1986): "Distinguishing outlier types in time series," *Journal of the Royal Statistical Society - Series B*, 48, 39–47. 98

Muirhead, R. (1982): *Aspects of multivariate statistical theory*, Whiley Series in Probability and Mathematical Statistics. 45, 81

Müller, U., M. Dacorogna, R. Dave, R. Olsen, R. Piquet, and J. von Weizsäcker (1997): "Volatilities of different time resolutions - Analyzing the dynamics of market components," *Journal of Empirical Finance*, 4, 213–239. 51

Müller, U. A., M. M. Dacorogna, R. Dave, O. V. Pictet, R. B. Olsen, and J. R. Ward (1993): "Fractals and intrinsic time - a challenge to econometricians," *XXXIX International AEA Conference on Real Time Econometris, 14-15 Oct 1993, Luxembourg*. 51

Nolan, J. (1997): "Maximum likelihood estimtiion of stable parameters," *unpublished manuscript.* 20

——— (2003): *Modeling financial data with stable distributions*, Handbook of heavily tailed distrutions in finance, Elsevier. 5, 6, 95

——— (2005): "Multivariate stable densities and distribution functions: general and elliptical case," *Deutsche Bundesank's 2005 Annual Fall Conference.* 7, 24

Ortobelli, S., I. Huber, and E. Schwartz (2002): "Portfolio selection with stable distributed returns," *Mathematical methods of operational research*, 55, 265–300. 4, 95

Ortobelli, S. and S. Rachev (2005): "Risk management and dynamic portfolio selection with stable Paretian," *mimeo.* 4, 95

Pagan, A. (1996): "The ecomometrics of financial markets," *Journal of Empirical Finance*, 3, 15–102. 4

Panorska, A., S. Mittnik, and S. Rachev (1995): "Stable GARCH models for financial time series," *Applied Mathematics Letters*, 7. 17, 95

Paolella, M. (1997): "Tail estimation and conditional modeling of heteroskedastic time-series," Ph.D. thesis, Istitute of Statistics and Econometrics, Christian Albrrechts University, Kiel, Germany. 34

——— (2001): "Testing the stable paretian assumption," *Mathematical and Computer Modelling*, 34, 1095–1112. 5, 16, 17, 34

Pong, S., M. Shackleton, S. Taylor, and X. Xu (2004): "Forecasting currency volatility: A comparison of implied volatilities and AR(FI)MA models," *Journal of Banking and Finance*, 28, 2514–2563. 51

Resnick, S. (1997): "Heavy tail modeling and teletraffic data," *Annals of Statistics*, 25, 1805–1869. 5

Rosenblatt, M. (1952): "Remarks on a multivariate transformation," *The Annals of Mathematical Statistics*, 23, 470–472. 70

Ruiz, E. and L. Pascual (2002): "Bootstrapping financial time series," *Journal of Economic Surveys*, 16, 271–300. 25

Samorodnitsky, G. and M. Taqqu (1994): *Stable non-gaussian random processes*, Chapman & Hall. 6, 8, 10, 124

Sentana, E. (1991): "Quadratic ARCH models: a potential re-interpretation of ARCH models," *Discussion paper, LSE Financial Market Group.* 16, 17

——— (1995): "Quadratic ARCH models," *Review of Economic Studies*, 62, 369–661. 16, 17

Sheppard, K. (2006): "Realized covariance and scrambling," *Manuscript.* 42, 59, 78

Thomakos, D. and T. Wang (2003): "Realized volatility in the futures markets," *Journal of Empirical Fianance*, 321–353. 51

TSE, Y. AND A. TSUI (2002): "A Multivariate Generalized Autoregressive Conditional Heteroscedasticity Model With Time-Varying Correlations," *Journal of Business and Economic Statistics*, 20, 351–362. 22

ZHANG, L. (2006): "Estimating covariation: Epps effect, microstructure noise." *Working Paper*. 42, 78

ZHANG, L., P. MYKLAND, AND Y. A. SAHALIA (2005): "A tale of two time scales: Determining integrated volatility with noisy high-frequency data," *Journal of the American Statistical Association*, 100, 1394–1411. viii, 91, 108

ZOLOTAREV, V. (1986): *One-dimensional stable distributions*, Translations of Mathematical Monographs, vol. 65. American Mathematical Society. 8